

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**ESCUELA DE POSGRADO  
MAESTRÍA EN CIENCIAS AMBIENTALES**



**“EVALUACIÓN DE PATRONES ESPACIALES Y  
PRONÓSTICO DE LA CALIDAD DEL AIRE EN LIMA  
METROPOLITANA MEDIANTE TÉCNICAS ESTADÍSTICAS  
MULTIVARIADAS”**

**Presentada por:  
JOSÉ ABEL ESPINOZA GUILLEN**

**TESIS PARA OPTAR EL GRADO DE MAESTRO  
MAGISTER SCIENTIAE EN CIENCIAS AMBIENTALES**

**Lima - Perú**

**2021**

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**ESCUELA DE POSGRADO  
MAESTRÍA EN CIENCIAS AMBIENTALES**

**“EVALUACIÓN DE PATRONES ESPACIALES Y  
PRONÓSTICO DE LA CALIDAD DEL AIRE EN LIMA  
METROPOLITANA MEDIANTE TÉCNICAS ESTADÍSTICAS  
MULTIVARIADAS”**

**TESIS PARA OPTAR EL GRADO DE MAESTRO  
MAGISTER SCIENTIAE**

**Presentada por:**

**JOSÉ ABEL ESPINOZA GUILLEN**

**Sustentada y aprobada ante el siguiente jurado:**

Mg.Sc. Jerónimo García Villanueva  
**PRESIDENTE**

Ph.D. Sergio Pacsi Valdivia  
**ASESOR**

Lic.Rer.Reg. Víctor Aguilar Vidangos  
**MIEMBRO**

Mg.Sc. Víctor Miyashiro Kiyari  
**MIEMBRO**

## DEDICATORIA

*El presente trabajo de investigación lo dedico  
a mis padres y hermanas que con su amor  
incondicional me impulsan a seguir  
mejorando día a día.*

*Y en memoria de mi profesora Juana  
Soledad Zavaleta Minaya, por iniciarme  
en el maravilloso mundo de la  
Ciencia, Tecnología e Innovación.*

## **AGRADECIMIENTOS**

A la Universidad Nacional Agraria La Molina, mi alma mater, la cual me brindo años fructíferos durante el pregrado y la maestría.

A mi patrocinador, Ph.D. Sergio Pacsi Valdivia por su amistad, enseñanza y continuo asesoramiento sin los cuales este trabajo de investigación no se hubiera cumplido.

A los señores miembros del jurado, los profesores Mg.Sc. Jerónimo García Villanueva, Arq. Víctor Aguilar Vidangos y Mg.Sc. Víctor Miyashiro Kiyon, por su orientación en los aspectos formales, conceptuales y metodológicos de este trabajo de investigación.

Al Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) por facilitarme la data de las concentraciones de contaminantes atmosféricos (particulado y gases) y variables meteorológicas de la Red de Monitoreo Automático de la Calidad del Aire de Lima Metropolitana.

A Marleni, mi compañera de vida, por todo el apoyo brindando en lo personal y profesional.

A mi familia, amigos de posgrado, pregrado, del colegio, amigos de toda la vida, por su apoyo y confianza. ¡gracias a todos!

## ÍNDICE GENERAL

I.	INTRODUCCIÓN .....	1
II.	REVISIÓN DE LITERATURA .....	5
	2.1 CONTAMINACIÓN ATMOSFÉRICA .....	5
	2.2 MATERIAL PARTICULADO (PM) .....	5
	2.3 DIÓXIDO DE AZUFRE (SO <sub>2</sub> ).....	7
	2.4 DIÓXIDO DE NITRÓGENO (NO <sub>2</sub> ).....	8
	2.5 OZONO TROPOSFÉRICO (O <sub>3</sub> ).....	9
	2.6 MONÓXIDO DE CARBONO (CO).....	10
	2.7 INVENTARIO DE EMISIONES.....	11
	2.7.1 Fuente móviles.....	11
	2.7.2 Fuentes fijas puntuales.....	12
	2.7.3 Fuentes fijas de área.....	13
	2.7.4 Emisiones globales.....	13
	2.8 ESTÁNDAR DE CALIDAD AMBIENTAL (ECA) PARA AIRE.	13
	2.9 REGLAMENTO DE LOS NIVELES DE ESTADOS DE ALERTA POR CONTAMINACIÓN DEL AIRE.....	15
	2.10 ÍNDICE DE CALIDAD DEL AIRE – INCA.....	15
	2.11 ANÁLISIS ESTADÍSTICO MULTIVARIADO .....	17
	2.12 TÉCNICAS ESTADÍSTICAS MULTIVARIADAS.....	18
	2.12.1 Análisis de componentes principales (ACP).....	18
	2.12.2 Análisis de clúster jerárquico (ACJ).....	19
	2.12.3 Regresión lineal múltiple (RLM).....	20
III.	MATERIALES Y MÉTODOS .....	21
	3.1 DESCRIPCIÓN DEL ÁREA DE ESTUDIO .....	21
	3.2 UBICACIÓN DE LAS ESTACIONES DE MONITOREO Y DATOS DISPONIBLES.....	22
	3.3 METODOLOGÍA .....	24
	3.3.1 Determinación de los patrones espaciales de los contaminantes atmosféricos (PM <sub>10</sub> , PM <sub>2,5</sub> , NO <sub>2</sub> , O <sub>3</sub> , CO y SO <sub>2</sub> ) medidos en la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de	24

	Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ).....	
3.3.2	Determinación de los patrones espaciales de las variables meteorológicas (temperatura de aire, humedad relativa, dirección y velocidad del viento y radiación solar) medidos en la red de estaciones meteorológicas automáticas del Área Metropolitana de Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster (ACJ).....	26
3.3.3	Pronóstico de las concentraciones de material particulado (PM <sub>10</sub> y PM <sub>2,5</sub> ) y gases (NO <sub>2</sub> , O <sub>3</sub> , CO y SO <sub>2</sub> ) en el área de Lima Metropolitana a través de la regresión lineal múltiple (RLM).....	27
IV.	RESULTADOS Y DISCUSIÓN .....	30
4.1	DETERMINACIÓN DE LOS PATRONES ESPACIALES DE LOS CONTAMINANTES ATMOSFÉRICOS MEDIDOS EN LA REMCA DEL ÁREA METROPOLITANA DE LIMA MEDIANTE EL ACP Y EL ACJ.....	30
4.1.1	Patrones espaciales de los contaminantes atmosféricos identificados a través del análisis de componentes principales (ACP).....	37
4.1.2	Patrones espaciales de los contaminantes atmosféricos identificados a través del análisis de clúster jerárquico (ACJ).....	46
4.2	DETERMINACIÓN DE LOS PATRONES ESPACIALES DE LAS VARIABLES METEOROLÓGICAS MEDIDOS EN LA RED DE ESTACIONES METEOROLÓGICAS AUTOMÁTICAS DEL ÁREA METROPOLITANA DE LIMA MEDIANTE EL ACP Y EL ACJ.....	52
4.2.1	Patrones espaciales de las variables meteorológicas identificados a través del análisis de componentes principales (ACP).....	61

4.2.2	Patrones espaciales de las variables meteorológicas identificados a través del análisis de clúster jerárquico (ACJ).....	70
4.3	PRONÓSTICO DE LAS CONCENTRACIONES DE MATERIAL PARTICULADO Y GASES EN EL ÁREA DE LIMA METROPOLITANA A TRAVÉS DE LA REGRESIÓN LINEAL MÚLTIPLE.....	74
4.3.1	Pronóstico del material particulado PM <sub>10</sub> .....	75
4.3.2	Pronóstico del material particulado PM <sub>2,5</sub> .....	77
4.3.3	Pronóstico del dióxido de nitrógeno (NO <sub>2</sub> ).....	80
4.3.4	Pronóstico del ozono troposférico (O <sub>3</sub> ).....	82
4.3.5	Pronóstico del monóxido de carbono (CO).....	84
4.3.6	Pronóstico del dióxido de azufre (SO <sub>2</sub> ).....	87
V.	CONCLUSIONES .....	92
VI.	RECOMENDACIONES .....	94
VII.	REFERENCIAS BIBLIOGRÁFICAS .....	95
VIII.	ANEXOS .....	104

## ÍNDICE DE TABLAS

Tabla 1.	Comparación de las propiedades básicas del PM: partículas finas (PM <sub>2,5</sub> ) versus partículas gruesas (PM <sub>10</sub> ).....	7
Tabla 2.	Emisiones estimadas de contaminantes del aire, por fuentes fijas puntuales de Lima-Callao (Ton/año - 2016) .....	12
Tabla 3.	Emisiones estimadas de contaminantes del aire, por fuentes fijas de área de Lima-Callao (Ton/año - 2016).....	13
Tabla 4.	Emisiones totales estimadas de contaminantes del aire, por tipo de fuente Lima-Callao (Ton/año - 2016).....	13
Tabla 5.	Estándares de Calidad Ambiental para Aire.....	14
Tabla 6.	Guías de Calidad del Aire de la OMS.....	14
Tabla 7.	Niveles de Estados de Alerta Nacionales para Contaminantes del Aire.....	15
Tabla 8.	Valores del Índice de Calidad del Aire.....	16
Tabla 9.	Cuidados y recomendaciones del Índice de Calidad del Aire .....	17
Tabla 10.	Índice de Calidad del Aire.....	17
Tabla 11.	Características generales de la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de Lima.....	23
Tabla 12.	Esquema del primer modelo empleado para el pronóstico de las concentraciones horarias de los contaminantes atmosféricos.....	28
Tabla 13.	Esquema del segundo modelo empleado para el pronóstico de las concentraciones horarias de los contaminantes atmosféricos.....	29
Tabla 14.	Estadísticas descriptivas de las concentraciones de los contaminantes atmosféricos.....	32
Tabla 15.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de PM <sub>10</sub> .....	37
Tabla 16.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de PM <sub>2,5</sub> .....	38
Tabla 17.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de NO <sub>2</sub> .....	38
Tabla 18.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de O <sub>3</sub> .....	39



Tabla 19.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de CO.....	40
Tabla 20.	Tabla: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de SO <sub>2</sub> .....	40
Tabla 21.	Pruebas de KMO y esfericidad de Bartlett de los contaminantes atmosféricos.....	41
Tabla 22.	Principales resultados de la aplicación del ACP a los datos de concentraciones de PM <sub>10</sub> .....	42
Tabla 23.	Principales resultados de la aplicación del ACP a los datos de concentraciones de PM <sub>2,5</sub> .....	42
Tabla 24.	Principales resultados de la aplicación del ACP a los datos de concentraciones de NO <sub>2</sub> .....	43
Tabla 25.	Principales resultados de la aplicación del ACP a los datos de concentraciones de O <sub>3</sub> .....	44
Tabla 26.	Principales resultados de la aplicación del ACP a los datos de concentraciones de CO.....	45
Tabla 27.	Principales resultados de la aplicación de ACP a los datos de concentraciones de SO <sub>2</sub> .....	46
Tabla 28.	Estadísticas descriptivas de las variables meteorológicas.....	53
Tabla 29.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de temperatura del aire.....	61
Tabla 30.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de humedad relativa .....	62
Tabla 31.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de dirección del viento.....	63
Tabla 32.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de velocidad del viento.....	63
Tabla 33.	Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de radiación solar .....	64
Tabla 34.	Pruebas de KMO y esfericidad de Bartlett de las variables meteorológicas .....	65
Tabla 35.	Principales resultados de la aplicación del ACP a los datos de temperatura del aire .....	65

Tabla 36.	Principales resultados de la aplicación del ACP a los datos de humedad relativa .....	67
Tabla 37.	Principales resultados de la aplicación del ACP a los datos de dirección del viento .....	67
Tabla 38.	Principales resultados de la aplicación del ACP a los datos de velocidad del viento .....	68
Tabla 39.	Principales resultados de la aplicación del ACP a los datos de radiación solar.....	69
Tabla 40.	Modelos de regresión lineal múltiple para el pronóstico horario de PM <sub>10</sub> e indicadores de desempeño .....	77
Tabla 41.	Modelos de regresión lineal múltiple para el pronóstico horario de PM <sub>2,5</sub> e indicadores de desempeño.....	79
Tabla 42.	Modelos de regresión lineal múltiple para el pronóstico horario de NO <sub>2</sub> e indicadores de desempeño.....	82
Tabla 43.	Modelos de regresión lineal múltiple para el pronóstico horario de O <sub>3</sub> e indicadores de desempeño.....	84
Tabla 44.	Modelos de regresión lineal múltiple para el pronóstico horario de CO e indicadores de desempeño.....	86
Tabla 45.	Modelos de regresión lineal múltiple para el pronóstico horario de SO <sub>2</sub> e indicadores de desempeño.....	89
Tabla 46.	Comparación con estudios similares de modelos de regresión lineal múltiple.....	89
Tabla 47.	Modelos de regresión lineal múltiple para el pronóstico horario de PM <sub>10</sub> e indicadores de desempeño para las estaciones VMT, CDM y PPD.....	90
Tabla 48.	Modelos de regresión lineal múltiple para el pronóstico horario de PM <sub>2,5</sub> e indicadores de desempeño para las estaciones VMT, CDM y PPD.....	91

## ÍNDICE DE FIGURAS

Figura 1.	Comparación de tamaños del $PM_{2,5}$ y $PM_{10}$ con el diámetro promedio de un cabello humano ( $\sim 70 \mu m$ ) y arena fina de playa ( $\sim 90 \mu m$ ).....	6
Figura 2.	Área de estudio y localización de las estaciones de la Red de Monitoreo Automático de la Calidad del Aire del Área Metropolitana de Lima.....	21
Figura 3.	Diagrama de cajas y bigotes para las concentraciones de $PM_{10}$ .....	33
Figura 4.	Diagrama de cajas y bigotes para las concentraciones de $PM_{2,5}$ .....	34
Figura 5.	Diagrama de cajas y bigotes para las concentraciones de $NO_2$ .....	35
Figura 6.	Diagrama de cajas y bigotes para las concentraciones de $O_3$ .....	35
Figura 7.	Diagrama de cajas y bigotes para las concentraciones de $CO$ .....	36
Figura 8.	Diagrama de cajas y bigotes para las concentraciones de $SO_2$ .....	36
Figura 9.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $PM_{10}$ .....	47
Figura 10.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $PM_{2,5}$ .....	48
Figura 11.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $NO_2$ .....	48
Figura 12.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $O_3$ .....	49
Figura 13.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $CO$ .....	50
Figura 14.	Dendrograma resultante de la aplicación del ACJ a las concentraciones de $SO_2$ .....	51
Figura 15.	Rosa de vientos para la estación ATE, periodo 2015-2019.....	54
Figura 16.	Rosa de vientos para la estación SBJ, periodo 2015-2019.....	55
Figura 17.	Rosa de vientos para la estación CDM, periodo 2015-2019.....	56
Figura 18.	Rosa de vientos para la estación STA, periodo 2015-2019.....	56
Figura 19.	Rosa de vientos para la estación VMT, periodo 2015-2019.....	57
Figura 20.	Rosa de vientos para la estación HCH, periodo 2015-2019.....	58
Figura 21.	Rosa de vientos para la estación SJL, periodo 2015-2019.....	58
Figura 22.	Rosa de vientos para la estación SMP, periodo 2015-2019.....	59

Figura 23.	Rosa de vientos para la estación CRB, periodo 2015-2019.....	60
Figura 24.	Rosa de vientos para la estación PPD periodo 2015-2019.....	60
Figura 25.	Dendrograma resultante de la aplicación del ACJ a los datos de temperatura del aire .....	70
Figura 26.	Dendrograma resultante de la aplicación del ACJ a los datos de humedad relativa .....	71
Figura 27.	Dendrograma resultante de la aplicación del ACJ a los datos de dirección del viento.....	72
Figura 28.	Dendrograma resultante de la aplicación del ACJ a los datos de velocidad del viento .....	73
Figura 29.	Dendrograma resultante de la aplicación del ACJ a los datos de radiación solar .....	74
Figura 30.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de PM <sub>10</sub> para la estación ATE .....	76
Figura 31.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de PM <sub>2,5</sub> para la estación ATE .....	79
Figura 32.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de NO <sub>2</sub> para la estación ATE .....	81
Figura 33.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de O <sub>3</sub> para la estación ATE.....	83
Figura 34.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de CO para la estación ATE .....	85
Figura 35.	Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de SO <sub>2</sub> para la estación ATE .....	88

## ÍNDICE DE ANEXOS

Anexo 1.	Salidas del modelo I para el pronóstico de las concentraciones horarias de contaminantes atmosféricos para la estación ATE.....	104
Anexo 2.	Salidas del modelo II para el pronóstico de las concentraciones horarias de contaminantes atmosféricos para la estación ATE.....	110
Anexo 3.	Salidas de los modelos I y II para el pronóstico de las concentraciones horarias de PM <sub>10</sub> y PM <sub>2,5</sub> para las estaciones VMT, CDM y PPD.....	116

## RESUMEN

En la presente investigación, se aplicaron los métodos estadísticos multivariados de análisis de componentes principales (ACP) y análisis de clúster jerárquico (ACJ) a los datos de concentraciones de seis contaminantes atmosféricos y cinco parámetros meteorológicos durante el periodo 2015-2019, con el objetivo de determinar los patrones espaciales de calidad de aire en el Área Metropolitana de Lima y Callao (AMLC). Asimismo, se realizó el pronóstico de las concentraciones de los contaminantes atmosféricos mediante modelos de regresión lineal múltiple (RLM). En el estudio se analizaron los valores de concentración horarias de PM<sub>10</sub>, PM<sub>2,5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO, SO<sub>2</sub> y variables meteorológicas como temperatura del aire, humedad relativa, dirección, velocidad del viento y radiación solar medidas en la Red de Monitoreo Automático de Calidad de Aire (REMCA), administrada por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). Los principales resultados mostraron que las diez estaciones de la REMCA se podrían agrupar según su comportamiento de contaminación del aire en seis grupos diferentes para el PM<sub>10</sub> y PM<sub>2,5</sub>, cinco grupos diferentes para NO<sub>2</sub> y CO, tres grupos diferentes para el O<sub>3</sub>, y siete grupos diferentes para el SO<sub>2</sub>. Así también, de acuerdo con la evaluación de las variables meteorológicas las estaciones se podrían agrupar en dos grupos diferentes para la temperatura del aire, humedad relativa y radiación solar; cinco grupos diferentes para la velocidad del viento; y siete grupos diferentes para la dirección del viento. Finalmente, los modelos implementados para el pronóstico de las concentraciones de los seis contaminantes de interés, presentaron mayor desempeño cuando se seleccionaron como variables predictores a las concentraciones del mismo contaminante de las otras estaciones de monitoreo. Asimismo, los coeficientes de determinación ajustados para todos los modelos fueron superiores al 41%. El modelo de pronóstico generado para el parámetro O<sub>3</sub> obtuvo los mejores indicadores de desempeño, llegando a presentar un  $r^2_{\text{ajustado}}$  de 0.86 con una RECM de 5.1  $\mu\text{g}/\text{m}^3$ .

**Palabras claves:** Contaminantes atmosféricos, Análisis estadístico multivariado, Análisis de componentes principales, Análisis de clúster jerárquico, Regresión lineal múltiple

## ABSTRACT

In this research, the multivariate statistical methods of principal component analysis (PCA) and hierarchical cluster analysis (HCA) were applied to the concentration data of six air pollutants and five meteorological parameters during the period 2015-2019 with the objective to determine the patterns spatial air quality in the Metropolitan Area of Lima and Callao (MALC). Likewise, the forecast of the concentrations of atmospheric pollutants was carried out using multiple linear regression models (MLR). The study analyzed the hourly concentration values of PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO, SO<sub>2</sub> and meteorological variables such as air temperature, relative humidity, direction, wind speed and solar radiation measured in the Automatic Monitoring Network were analyzed. of Air Quality (AAQMN), administered by the National Service of Meteorology and Hydrology of Peru (SENAMHI). The main results showed that the ten AAQMN stations could be grouped according to their air pollution behavior into six different groups for PM<sub>10</sub> and PM<sub>2.5</sub>, five different groups for NO<sub>2</sub> and CO, three different groups for O<sub>3</sub>, and seven different groups for SO<sub>2</sub>. Also, according to the evaluation of the meteorological variables, the stations could be grouped into two different groups for air temperature, relative humidity and solar radiation; five different groups for wind speed; and seven different groups for the wind direction. Finally, the models implemented for the forecasting of the concentrations of the six pollutants of interest presented higher performance when the concentrations of the same pollutant from the other monitoring stations were selected as predictor variables. Likewise, the coefficients of determination adjusted for all models were greater than 41%. The forecast model generated for the O<sub>3</sub> parameter had the best performance indicators, reaching an adjusted r-squared of 0.86 with a RMSE of 5.1 µg/m<sup>3</sup>.

**Keywords:** Air Pollutants, Multivariate Statistical Analysis, Principal Component Analysis, Hierarchical Cluster Analysis, Multiple Linear Regression

## I. INTRODUCCIÓN

El Área Metropolitana de Lima – Callao (AMLC) es una de las mega ciudades en América Latina, la cual presenta altos niveles de contaminantes atmosféricos, especialmente material particulado (PM). Los niveles de material particulado con diámetro menor a 10 micras ( $PM_{10}$ ) y material particulado con diámetro menor a 2,5 micras ( $PM_{2,5}$ ) en ciertas zonas de la ciudad a menudo exceden las guías de calidad del aire de la Organización Mundial de la Salud (OMS) y los Estándares de Calidad Ambiental (ECA) para Aire, aprobados por el Ministerio del Ambiente (MINAM) (Silva *et al.* 2017, 2018).

Para controlar y mejorar la calidad del aire, el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) ha establecido diez estaciones de monitoreo, que conforman la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del AMLC. Estas estaciones se distribuyen en diferentes zonas del AMLC y están equipadas con varios instrumentos para medir los contaminantes atmosféricos. Algunas de ellas están ubicadas físicamente muy cercanas o se caracterizan por una condición de monitoreo similar, que generalmente presentan comportamientos semejantes (Sánchez *et al.* 2015, Sánchez y Ordoñez 2016, Silva *et al.* 2017).

El uso de redes de monitoreo asociadas con los recursos de modelación computacional es fundamental para monitorear la evolución de las dispersiones de contaminantes y así obtener soporte para la elaboración de políticas públicas ambientales (Silva *et al.* 2017, 2018). Por lo tanto, es necesario y significativo evaluar los patrones espaciales de calidad del aire y pronosticar las concentraciones de material particulado y gases, utilizando métodos estadísticos multivariados, como el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ) (Pires *et al.* 2008, Azid *et al.* 2015, Wang *et al.* 2018).

Los métodos estadísticos como el ACP y el ACJ se han utilizado en muchos estudios destinados a la gestión de la calidad del aire, optimizando redes de monitoreo automáticos e identificando fuentes de contaminación atmosférica. En estudios anteriores, estos enfoques se combinan para explorar información significativa a partir de los datos de origen. Gramsch



*et al.* (2006) utilizaron el ACP y el ACJ para determinar las tendencias estacionales y la distribución espacial de PM<sub>10</sub> y O<sub>3</sub> en Santiago de Chile, concluyendo que la ciudad tenía cuatro grandes sectores con comportamientos de contaminación del aire diferentes.

Shah y Shaheen (2008) emplearon estas técnicas para identificar la principal fuente de metales traza en el aire en el área de Islamabad. Pires *et al.* (2008a, b) los aplicaron en las concentraciones máxicas de SO<sub>2</sub>, PM<sub>10</sub>, CO, NO<sub>2</sub> y O<sub>3</sub> para el manejo eficiente de la calidad del aire en el Área Metropolitana de Oporto. Azid *et al.* (2015a, b) los utilizaron para investigar los patrones espaciales de la calidad del aire e identificar los parámetros más significativos en Malasia.

El INECC (2018) realizó una evaluación de redundancia de las estaciones de monitoreo de calidad de aire en el estado de México utilizando el ACP y el ACJ, estas técnicas se aplicaron para cada contaminante CO, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2,5</sub> y SO<sub>2</sub>, esto permitió identificar y verificar las estaciones que presentan comportamientos similares entre sí. Núñez *et al.* (2019) aplicaron las mismas técnicas en los contaminantes NO<sub>x</sub> (NO + NO<sub>2</sub>), O<sub>3</sub> y PM<sub>10</sub> en la ciudad y provincia de Madrid, los resultados obtenidos permitieron la clasificación de diferentes estaciones de monitoreo en base a cada uno de los contaminantes evaluados. Los estudios antes mencionados, verificaron que estos métodos pueden complementarse entre sí y la combinación de ellos podría proporcionar un enfoque alternativo práctico para analizar y resolver problemas ambientales.

Sin embargo, en el AMLC, hay una ausencia de estudios y modelos que permiten determinar los patrones espaciales y predecir las concentraciones de contaminantes atmosféricos como el material particulado y gases. Por lo tanto, este trabajo de investigación evaluó los patrones espaciales de calidad del aire en el Área Metropolitana de Lima, en las diez estaciones de monitoreo de la REMCA, con el fin de identificar áreas con comportamiento similar de contaminantes en la ciudad de Lima.

La contaminación atmosférica tiene efectos adversos sobre el ambiente y el estado de salud de una población determinada. Esto se debe a que, el aire está contaminado por diversas fuentes y su complejidad depende en gran medida de las características meteorológicas, la topografía y la naturaleza de los contaminantes emitidos. El rápido avance en el nivel de las actividades industriales y el crecimiento de la población en la segunda mitad del siglo XX representa una seria amenaza para la calidad del aire y el bienestar humano. Además, muchas

descargas accidentales de contaminantes atmosféricos provocan una exposición aguda y un riesgo potencial para la salud humana; el Área Metropolitana de Lima no es ajena a estos problemas (Azid *et al.* 2015, Pacsi 2016, OMS 2018).

A pesar del esfuerzo realizado por los gobiernos y las partes interesadas para monitorear y controlar el nivel de contaminación del aire, su impacto negativo y peligroso es enorme. Sin embargo, una estimación de la variabilidad espacial de la contaminación del aire puede entenderse mejor si se tiene en cuenta la ubicación de los sitios de seguimiento dentro de un área de estudio. Según estimaciones del 2016, la contaminación atmosférica causa al año 4.2 millones de muertes prematuras en las ciudades y zonas rurales del todo el mundo, siendo el 91 por ciento de estas muertes producidas en países de ingresos bajos y medianos, por otro lado, las mayores tasas de enfermedades se registran en las regiones del Pacífico Occidental y Asia Sudoriental (OMS, 2018).

En un estudio realizado por la Organización Mundial de la Salud en el año 2016 en el que se monitorearon a 3000 ciudades en 103 países, se identificó a Lima como una de las ciudades más contaminadas a nivel de América Latina por  $PM_{10}$  y  $PM_{2,5}$ , siendo sus pobladores expuestos a valores de contaminación que exceden las Directrices de la OMS sobre Calidad del Aire (OMS, 2016). Monitorear la calidad del aire es esencial para el control de la contaminación atmosférica.

El monitoreo de la calidad del aire implica la observación de grandes conjuntos de datos complejos en estaciones, que requieren la integración de técnicas estadísticas modernas y sólidas para simplificar y evitar malas interpretaciones. Los datos ambientales generados se deben utilizar principalmente para análisis estadísticos en los que se basen las estrategias para identificar las principales fuentes contribuyentes de contaminación en el sitio monitoreado (Pires *et al.* 2008, Moura *et al.* 2018, Núñez *et al.* 2019).

Algunas de las principales herramientas estadísticas utilizadas para ayudar e interpretar los datos de monitoreo de la calidad del aire son el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ) (Ibarra *et al.* 2009). Estas técnicas estadísticas multivariadas se han aplicado con éxito en la gestión ambiental de varias partes del mundo; sin embargo, en el Perú son técnicas incipientes. Los pocos estudios realizados se han dado en la gestión de la calidad del recurso hídrico en los ríos Rímac, Jequetepeque y Utcubamba (Espíritu 2010, Flores J. 2017, Flores R. 2017, Corroto *et al.* 2018); mientras que, en la

gestión de la calidad del aire solo se conoce de un estudio pionero realizado por el SENAMHI en el año 2013, en el que se utilizó el análisis factorial (AF) para la determinación de las principales fuentes de contaminación del aire (SENAMHI 2016).

Es por todo lo expuesto anteriormente que se necesita evaluar los patrones espaciales de calidad del aire e identificar áreas en Lima Metropolitana con comportamientos similares de contaminación del aire mediante las técnicas multivariadas de ACP y ACJ; y pronosticar las concentraciones de material particulado y gases, con fines de insumo para los tomadores de decisiones en la prevención y control de la contaminación del aire.

El presente trabajo de investigación tuvo el siguiente objetivo general:

- Evaluar los patrones espaciales de calidad del aire y pronosticar las concentraciones de material particulado y gases en el Área Metropolitana de Lima mediante la aplicación de técnicas estadísticas multivariadas.

El cual a su vez desencadenó, los siguientes objetivos específicos:

- Determinar los patrones espaciales de los contaminantes atmosféricos ( $PM_{10}$ ,  $PM_{2,5}$ ,  $NO_2$ ,  $O_3$ ,  $CO$  y  $SO_2$ ) medidos en la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ).
- Determinar los patrones espaciales de las variables meteorológicas (temperatura de aire, humedad relativa, dirección y velocidad del viento y radiación solar) medidos en la red de estaciones meteorológicas automáticas del Área Metropolitana de Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ).
- Pronosticar las concentraciones de material particulado ( $PM_{10}$  y  $PM_{2,5}$ ) y gases ( $NO_2$ ,  $O_3$ ,  $CO$  y  $SO_2$ ) en el área de Lima Metropolitana a través de la regresión lineal múltiple (RLM).

## II. REVISIÓN DE LITERATURA

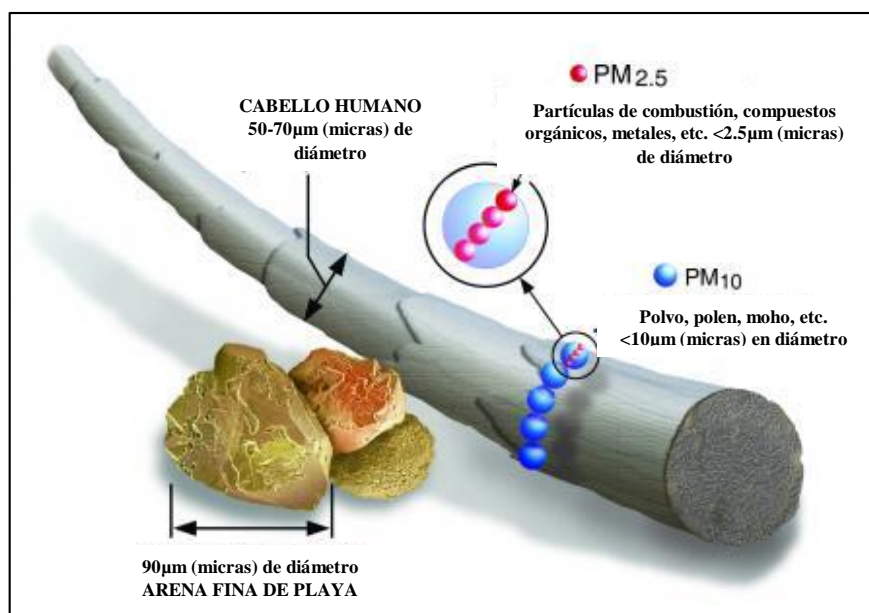
### 2.1 CONTAMINACIÓN ATMOSFÉRICA

La contaminación atmosférica es un proceso que introduce en la atmósfera diversos agentes que causan daño a los seres humanos, a otros organismos vivos y al ambiente en general (Kim *et al.* 2015). Wark y Warner (2007) lo definieron como la presencia en la atmósfera de uno o más contaminantes o sus combinaciones, en cantidades tales y con tal duración que pueden afectar la vida de las personas, animales y plantas, e interferir con el goce de la vida, la propiedad y el ejercicio de las actividades.

### 2.2 MATERIAL PARTICULADO (PM)

El material particulado es una mezcla heterogénea de partículas sólidas y gotas líquidas que se encuentran suspendidas en el aire ambiental que varían continuamente su tamaño y composición en el espacio y tiempo (WHO 2018). Entre sus constituyentes tenemos a: nitratos, sulfatos, carbono elemental y orgánico, compuestos orgánicos (por ejemplo, hidrocarburos aromáticos policíclicos, conocidos como HAPs), compuestos biológicos (por ejemplo, endotoxinas, fragmentos celulares) y metales (por ejemplo, hierro, cobre, níquel, zinc y vanadio) (Kim *et al.* 2015, WHO 2018).

Existen diferentes formas de clasificar al material particulado, no obstante, el diámetro aerodinámico es uno de los principales criterios para describir su capacidad de transporte en la atmósfera y/o capacidad de inhalación a través del sistema respiratorio del organismo (Pires *et al.* 2008, WHO 2018). La Agencia de Protección Ambiental de los Estados Unidos (USEPA) ha estado regulando partículas principalmente en dos categorías de tamaño, basadas en su capacidad de penetración en el pulmón, es así que, tenemos a las partículas gruesas (PM<sub>10</sub>) con un diámetro aerodinámico igual o menor a 10 µm y las partículas finas (PM<sub>2,5</sub>) con diámetro aerodinámico menor o igual a 2.5 µm (USEPA 2020). En la **Figura 1** se muestra una comparación de tamaños entre el PM<sub>2,5</sub> y PM<sub>10</sub> frente al diámetro medio de un cabello humano (~ 70 µm) y arena fina de playa (~ 90 µm).



**Figura 1: Comparación de tamaños del PM<sub>2,5</sub> y PM<sub>10</sub> con el diámetro promedio de un cabello humano (~ 70 µm) y arena fina de playa (~ 90 µm)**

FUENTE: USEPA 2020.

Las fuentes de material particulado pueden explicarse como la emisión directa al aire o como conversión de precursores de gases, tales como dióxido de azufre, óxidos de nitrógeno, amoníaco y compuestos orgánicos volátiles no metánicos (COVNM), los cuales son liberados tanto de fuentes antropogénicas como naturales (WHO 2006, 2018). Las fuentes antropogénicas son muy variables e incluyen la combustión de combustibles sólidos (carbón, lignito, petróleo pesado y biomasa), las actividades industriales y agrícolas, la erosión del pavimento por el tráfico por carretera y la abrasión de frenos y neumáticos (Srimuruganandam y Nagendra, 2012). Por el contrario, los que ocurren naturalmente pueden ser explicados por fuentes como volcanes, tormentas de polvo, incendios forestales, vegetación viva y rocío (WHO 2006, USEPA 2020). En la **Tabla 1** se presenta una comparación de las propiedades generales que distinguen a las partículas finas (PM<sub>2,5</sub>) y las gruesas (PM<sub>10</sub>).

El tráfico por carreteras es una de las principales fuentes de material particulado, originada principalmente por el desgaste de componentes del vehículo, tales como frenos y neumáticos, así como la suspensión del polvo de la carretera (Kim *et al.* 2015). Las partículas inorgánicas de material cristal procedentes de la abrasión del pavimento son a menudo ricas en minerales que contienen silicio (Si), aluminio (Al), potasio (K), sodio (Na) y calcio (Ca), mientras las partículas de desgaste de frenos y neumáticos pueden contener metales como

cobre (Cu), antimonio (Sb), plomo (Pb), cadmio (Cd) y zinc (Zn) (Srimuruganandam y Nagendra, 2012).

**Tabla 1: Comparación de las características básicas del PM: partículas finas (PM<sub>2,5</sub>) versus partículas gruesas (PM<sub>10</sub>)**

Características	Partículas finas (PM <sub>2,5</sub> )	Partículas gruesas (PM <sub>10</sub> )
Diámetro	Menos de 2.5 µm	Menos de 10 µm
Composición	Sulfato, SO <sub>4</sub> <sup>-2</sup> ; Nitrato, NO <sub>3</sub> <sup>-</sup> ; Amonio, NH <sub>4</sub> <sup>+</sup> ; Hidrógeno, H <sup>+</sup> ; Carbono elemental, C; compuestos orgánicos; PAH; Metales, Pb, Cd, V, Ni, Cu, Zn; Agua unida a partículas; y orgánicos biogénicos.	Polvo resuspendido, polvo del suelo, polvo de la calle; carbón y cenizas volantes; óxidos metálicos de Si, Al, Mg, Ti, Fe, CaCO <sub>3</sub> , NaCl, sal marina; polen, esporas de moho y partes de plantas.
Fuentes	Combustión de carbón, petróleo, gasolina; Productos de transformación de NO <sub>x</sub> , SO <sub>2</sub> y compuestos orgánicos, incluyendo orgánicos biogénicos, por ejemplo, terpenos; Procesos de alta temperatura; Fundiciones y acerías.	Resuspensión del suelo rastreado en carreteras y calles; Suspensión de suelos perturbados, por ejemplo, agricultura, minería; Resuspensión de polvos industriales; Construcción, combustión del carbón y del petróleo, y rocío del océano.
Tiempo de vida	Días a semanas	Minutos a horas
Distancia de viaje (Kilómetros)	100 a 1000	1 a 10

FUENTE: Srimuruganandam y Nagendra 2012, Kim *et al.* 2015.

### 2.3 DIÓXIDO DE AZUFRE (SO<sub>2</sub>)

El SO<sub>2</sub> es uno de los contaminantes más importantes, ya que resulta de la combustión de compuestos de azufre. Los volcanes y los océanos son las principales fuentes naturales de SO<sub>2</sub> (WHO 2018, USEPA 2019). Las emisiones antropogénicas de SO<sub>2</sub> provienen de la combustión de combustibles fósiles (principalmente carbón y aceites pesados), la quema de biomasa y la fundición de minerales que contienen azufre. Se han hecho muchos esfuerzos para reducir las emisiones de SO<sub>2</sub>; en consecuencia, en los últimos 30 años los niveles atmosféricos de SO<sub>2</sub> han ido disminuyendo continuamente en la mayoría de países industrializados occidentales (Aas *et al.* 2019).

El SO<sub>2</sub> y sus subproductos de oxidación se eliminan de la atmósfera por deposición húmeda y seca (precipitación ácida). Además, de estos procesos de transformación y eliminación, el SO<sub>2</sub> puede llevarse a cabo a grandes distancias, causando contaminación transfronteriza (Pires *et al.* 2008).

El SO<sub>2</sub> es un gas irritante cuando se inhala y en altas concentraciones, puede causar dificultades respiratorias en personas directamente expuestas a él. La absorción de SO<sub>2</sub> en las membranas mucosas de la nariz y en el tracto respiratorio superior resulta de su solubilidad en agua (WHO 2006). Los efectos de la inhalación de SO<sub>2</sub> aparecen en solo unos minutos, y las personas que sufren de asma y enfermedad pulmonar crónica, pueden ser especialmente susceptibles a sus efectos adversos (WHO 2018).

Este contaminante también afecta a las plantas y, dependiendo de sus niveles de concentración de masa, puede producir: degradación de la clorofila; reducción de la fotosíntesis; aumento de las tasas de respiración; y cambios en el metabolismo de las proteínas (USEPA 2019). Sin embargo, el azufre es un nutriente importante para las plantas debido al hecho que, por ejemplo, el azufre atmosférico puede ser absorbido por las hojas de algunas especies, lo que contribuye a la vitalidad de la planta en suelos con bajas concentraciones de azufre, como los suelos calcáreos (Pires *et al.* 2008).

El SO<sub>2</sub> combinado con otros contaminantes del aire y en condiciones específicas de humedad relativa, temperatura y precipitación; es responsable del deterioro de materiales, como metales y piedras. En muchas partes de Europa, algunos monumentos que resistieron el deterioro durante cientos o incluso miles de años han mostrado una degradación acelerada de su superficie en las últimas décadas (Pires *et al.* 2008).

## **2.4 DIÓXIDO DE NITRÓGENO (NO<sub>2</sub>)**

Los óxidos de nitrógeno (NO<sub>x</sub>) son un grupo de gases altamente reactivos. Estos gases contienen átomos de nitrógeno y oxígeno en diferentes proporciones. El óxido de nitrógeno (NO) y el NO<sub>2</sub> son los gases más importantes de este grupo y son considerados contaminantes significativos en la troposfera (Malley *et al.* 2018, WHO 2018).

Las emisiones antropogénicas de NO<sub>x</sub> resultan de los procesos de combustión, incluyendo los automóviles, las centrales eléctricas y otras fuentes industriales, comerciales y residenciales que queman combustibles fósiles. Los eventos naturales, como los procesos

biológicos anaerobios en el suelo y el agua, la actividad volcánica y la destrucción fotoquímica de los compuestos de nitrógeno en la atmósfera superior, también tienen una gran contribución en las emisiones de NO<sub>x</sub> (Habeebullah 2015, WHO 2018, USEPA 2019).

Los NO<sub>x</sub> tienen diversos efectos negativos (Pires *et al.* 2008, WHO 2018), como:

- la formación de lluvia ácida: reaccionan con otras sustancias en el aire para formar ácidos que caen a la tierra como lluvia, niebla, nieve o partículas secas;
- el deterioro de la calidad del agua: el aumento de la emisión de nitrógeno en los cuerpos de agua, particularmente en los estuarios costeros, alterando el equilibrio químico de los nutrientes utilizados por las plantas y animales acuáticos;
- la formación de productos químicos tóxicos: los NO<sub>x</sub> reaccionan con productos químicos orgánicos comunes para formar una amplia variedad de productos tóxicos, como radicales nitratos, nitroarenos y nitrosaminas;
- la reducción de la visibilidad: el NO<sub>2</sub> puede bloquear la transmisión de la luz;
- la contribución a un aumento de la temperatura de la tierra: uno de los NO<sub>x</sub>, el óxido nitroso (N<sub>2</sub>O), es un gas de efecto invernadero; y
- la formación de O<sub>3</sub> a nivel del suelo: en presencia de hidrocarburos y luz solar, los NO<sub>x</sub> contribuyen a la formación de O<sub>3</sub> troposférico, que puede causar problemas respiratorios graves.

## **2.5 OZONO TROPOSFÉRICO (O<sub>3</sub>)**

El O<sub>3</sub> es un fuerte oxidante fotoquímico que se encuentra en la troposfera y en otras capas de la atmósfera. El O<sub>3</sub> tiene un papel importante en la estratosfera (protección contra la radiación ultravioleta), sin embargo, en la tropósfera esta molécula es irritante y reactiva teniendo impactos negativos en la salud humana, el clima, la vegetación y los materiales (WHO 2006, Pires *et al.* 2008).

Con respecto a los efectos sobre la salud humana, los más importantes son:

- el daño a los tejidos del tracto respiratorio;
- la muerte de las células pulmonares y aumento de las tasas de replicación celular (hiperplasia);
- la inflamación de las vías aéreas; y



- el aumento de los síntomas respiratorios, como tos, dolor torácico (subesternal), dificultad para respirar profundamente y, en algunos casos, dolores de cabeza o náuseas (WHO 2006).

Con respecto al clima, se espera que un aumento de temperatura esté relacionado con el aumento de O<sub>3</sub> troposférico, porque es un gas de efecto invernadero e influye en el tiempo de residencia atmosférica de otros gases de efecto invernadero (Pires *et al.* 2008, Munir *et al.* 2015).

En la vegetación, causa daños en las hojas, en el crecimiento y en la reducción del rendimiento; así también, cambios en la sensibilidad por estrés biótico y abiótico (Pires *et al.* 2008, USEPA 2020).

Con respecto a los materiales, el O<sub>3</sub> en combinación con otros contaminantes atmosféricos, contribuyen a aumentar la corrosión de los materiales de construcción como el acero, el zinc, el cobre, el aluminio y el bronce (Pires *et al.* 2008, WHO 2018, USEPA 2020).

La presencia de O<sub>3</sub> en la troposfera es el resultado de tres procesos básicos:

- la producción fotoquímica por la interacción de hidrocarburos y NO<sub>x</sub> (emitidos por vapores de gasolina, centrales eléctricas de combustibles fósiles, refinerías y otras industrias) bajo la acción de las condiciones meteorológicas ambientales adecuadas (WHO 2006, USEPA 2020);
- el intercambio troposférico/estratosférico que provoca el transporte de aire estratosférico, rico en O<sub>3</sub>, hacia la troposfera (Pires *et al.* 2008); y
- el transporte horizontal debido al viento que trae O<sub>3</sub> producido en otras regiones (Pires *et al.* 2008, Munir *et al.* 2015).

## **2.6 MONÓXIDO DE CARBONO (CO)**

El CO es un gas incoloro, prácticamente inodoro, insípido y no irritante que es el resultado de la oxidación incompleta del carbono durante el proceso de la combustión (WHO 2018).

Las fuentes principales son la combustión incompleta de combustible (que ocurre cuando la relación aire-combustible presenta valores bajos), las emisiones industriales y otras fuentes de combustión (como carbón, gas, madera y queroseno). Las fuentes naturales, como los

volcanes, los gases naturales en las minas de carbón y los incendios forestales, también tienen una contribución importante (Pires *et al.* 2008, USEPA 2019).

Las emisiones de CO aumentan significativamente durante el clima frío. En estas condiciones, los motores necesitan más combustible para funcionar, y algunos dispositivos de control de emisiones, como sensores de oxígeno y convertidores catalíticos, funcionan de manera menos eficiente cuando están fríos (Rozante *et al.* 2017, USEPA 2019).

Los efectos sobre la salud humana, asociados con la presencia de CO dependen de su concentración y la duración de la exposición. A bajas concentraciones, el CO causa fatiga en personas sanas y dolor en el pecho en personas con enfermedades cardíacas. A altas concentraciones, causa dolores de cabeza, confusión y náuseas (WHO 2018).

Cuando se inhala, el CO se absorbe desde los pulmones hacia el torrente sanguíneo, donde forma un complejo con la hemoglobina (Hb), llamada carboxihemoglobina (COHb). La presencia de este complejo reduce la capacidad de transporte de oxígeno, causando hipoxia (bajo nivel de oxígeno disponible para los tejidos del cuerpo) (WHO 2006, 2018). El CO es uno de los contaminantes encontrados normalmente en la atmósfera que requiere medidas de prevención y control para garantizar una protección adecuada de la salud humana (Pires *et al.* 2008, USEPA 2019).

## **2.7 INVENTARIO DE EMISIONES**

La Comisión Multisectorial para la Gestión de la Iniciativa del Aire Limpio para Lima y Callao, presidida por el Ministerio del Ambiente (MINAM), elaboró el último inventario de emisiones para Lima y Callao publicado en el año 2019, tomando como año base el 2016 (MINAM 2019). Los principales resultados son los siguientes:

### **2.7.1 Fuente móviles**

Del total de emisiones vehiculares de PM<sub>2,5</sub> (3 897 ton/año), los vehículos a diésel son los que mayor aportan con un 97 por ciento; mientras que, la categoría vehicular que más contribuye son los ómnibus, lo cual representa el 42 por ciento del total de emisiones vehiculares. Así también, los vehículos con tecnología pre-euro son los que mayor aportan a las emisiones vehiculares totales de este contaminante con un 65 por ciento.

En el caso de los NO<sub>x</sub>, del total de emisiones de fuentes móviles (152 106 ton/año), el 89 por ciento se debe a la combustión de vehículos a diésel; siendo la categoría vehicular de los ómnibus, la que mayor contribuye a las emisiones vehiculares totales de este contaminante con un 42 por ciento.

En el caso del CO, del total de emisiones de fuentes móviles (272 905 ton/año), el 87 por ciento se debe a la combustión de vehículos a gasolina y gasohol; siendo la categoría vehicular de las motos, la que más contribuye con hasta un 50 por ciento, a las emisiones vehiculares totales de este contaminante.

Del total de emisiones vehiculares de SO<sub>2</sub> (38 764 ton/año), la combustión de los vehículos a gasolina/gasohol son la principal fuente, ya que aportan con un 95 por ciento a las emisiones vehiculares totales de este contaminante.

### 2.7.2 Fuentes fijas puntuales

Las emisiones estimadas de contaminantes atmosféricos emitidas por fuentes fijas puntuales se presentan en la **Tabla 2**. Los resultados muestran que, la principal fuente de emisión de PM<sub>2,5</sub>, PM<sub>10</sub> y SO<sub>2</sub>, es la fundición de Zinc de Cajamarquilla, con un aporte de 91.9, 92.8 y 73.1 por ciento, respectivamente. Mientras que, la refinación de derivados de petróleo es la fuente que contribuye en mayor cantidad al contaminante CO, actividad que se ejecuta en la Refinería La Pampilla, administrada por Repsol; y en la Refinería Conchán, administrada por Petroperú (MINAM 2019).

**Tabla 2: Emisiones estimadas de contaminantes del aire, por fuentes fijas puntuales de Lima-Callao (Ton/año - 2016)**

Fuentes	PM <sub>2,5</sub>	%	PM <sub>10</sub>	%	NO <sub>x</sub>	%	SO <sub>2</sub>	%	CO	%
Industria de Cemento (Hornos)	7.5	0.4	13.5	0.6	-	-	864.8	3.7	-	-
Harina de Pescado	111.1	6.2	111.1	5.2	-	-	-	-	-	-
Refinación de petróleo crudo	-	-	-	-	307.5	100	4715.0	20.2	461.3	100
Refinación de Cobre	25.8	1.4	30.8	1.4	-	-	709.6	3.0	-	-
Refinación de Zinc	1639.1	91.9	2001.1	92.8	-	-	17074.0	73.1	-	-
<b>Total</b>	<b>1783.5</b>	<b>100</b>	<b>2156.4</b>	<b>100</b>	<b>307.5</b>	<b>100</b>	<b>23363.3</b>	<b>100.0</b>	<b>461.3</b>	<b>100</b>

Nota: “-”, corresponde a un valor no estimado debido a que no es considerado por la referencia bibliográfica que establece los factores de emisión para determinado rubro industrial.

FUENTE: MINAM (2019)

### 2.7.3 Fuentes fijas de área

Las estimaciones de emisiones de las fuentes de área fija priorizadas para Lima y Callao, muestran que, la principal fuente que contribuye a las emisiones de PM<sub>2,5</sub>, PM<sub>10</sub>, NO<sub>x</sub>, SO<sub>2</sub> y CO son los restaurantes-pollerías, tal como se observa en la **Tabla 3** (MINAM 2019).

**Tabla 3: Emisiones estimadas de contaminantes del aire, por fuentes fijas de área de Lima-Callao (Ton/año - 2016)**

Fuentes	PM <sub>2,5</sub>	PM <sub>10</sub>	NO <sub>x</sub>	SO <sub>2</sub>	CO
Restaurantes -Pollerías	1107.7	1200.0	1774.4	8615.5	9548.8
<b>Total</b>	<b>1107.7</b>	<b>1200.0</b>	<b>1774.4</b>	<b>8615.5</b>	<b>9548.8</b>

Nota: “-”, corresponde a un valor no estimado debido a que no es considerado por la referencia bibliográfica que establece los factores de emisión para determinado rubro industrial.

FUENTE: MINAM (2019)

### 2.7.4 Emisiones globales

Se presentan los resultados totalizados de todas las fuentes móviles y fijas, priorizadas para Lima y Callao. Las fuentes móviles son las más importantes dentro de todas las fuentes de emisión, superando más del 50 por ciento. Estas aportan con el 57.4, 53.7, 98.6, 54.8 y 96.5 por ciento, a las emisiones de PM<sub>2,5</sub>, PM<sub>10</sub>, NO<sub>x</sub>, SO<sub>2</sub> y CO, respectivamente (**Tabla 4**) (MINAM 2019).

**Tabla 4: Emisiones totales estimadas de contaminantes del aire, por tipo de fuente Lima-Callao (Ton/año - 2016)**

Tipos de fuentes	PM <sub>2,5</sub>	%	PM <sub>10</sub>	%	NO <sub>x</sub>	%	SO <sub>2</sub>	%	CO	%
Móviles	3897.6	57.4	3897.6	53.7	152106.1	98.6	38764.5	54.8	272905.1	96.5
Puntuales	1783.5	26.3	2156.4	29.7	307.5	0.2	23363.3	33.0	461.3	0.2
Área	1107.7	16.3	1200.0	16.5	1774.4	1.2	8615.5	12.2	9548.8	3.4
<b>Total</b>	<b>6788.8</b>	<b>100</b>	<b>7254.0</b>	<b>100</b>	<b>154188.0</b>	<b>100.0</b>	<b>70743.3</b>	<b>100</b>	<b>282915.2</b>	<b>100</b>

FUENTE: MINAM (2019)

## 2.8 ESTÁNDAR DE CALIDAD AMBIENTAL (ECA) PARA AIRE

Según la Ley General del Ambiente N° 28611, el Estándar de Calidad Ambiental (ECA) es la medida que establece el nivel de concentración o del grado de elementos, sustancias o parámetros físicos, químicos y biológicos, presentes en el aire, agua o suelo, en su condición de cuerpo receptor, que no representa riesgo significativo para la salud de las personas ni al ambiente. Son expresados en máximos, mínimos o rangos.

**Tabla 5: Estándares de Calidad Ambiental para Aire**

Parámetros	Periodo	Valor [ $\mu\text{g}/\text{m}^3$ ]	Criterios de evaluación	Método de análisis <sup>[1]</sup>
Dióxido de Azufre (SO <sub>2</sub> )	24 horas	250	NE más de 7 veces al año	Fluorescencia ultravioleta (Método automático)
Dióxido de Nitrógeno (NO <sub>2</sub> )	1 hora	200	NE más de 24 veces al año	Quimioluminiscencia (Método automático)
	Anual	100	Media aritmética anual	
Material Particulado con diámetro menor a 2,5 micras (PM <sub>2,5</sub> )	24 horas	50	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	25	Media aritmética anual	
Material Particulado con diámetro menor a 10 micras (PM <sub>10</sub> )	24 horas	100	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	50	Media aritmética anual	
Monóxido de Carbono (CO)	1 hora	30000	NE más de 1 vez al año	Infrarrojo no dispersivo (NDIR) (Método automático)
	8 horas	10000	Media aritmética móvil	
Ozono (O <sub>3</sub> )	8 horas	100	Máxima media diaria NE más de 24 veces al año	Fotometría de absorción ultravioleta (Método automático)

NE: No Exceder. <sup>[1]</sup> o método equivalente aprobado.

FUENTE: Decreto Supremo N° 003-2017-MINAM

**Tabla 6: Guías de Calidad del Aire de la OMS**

Parámetro	Periodo	Valor [ $\mu\text{g}/\text{m}^3$ ]
SO <sub>2</sub>	media de 24 horas	20
	media de 10 minutos	500
NO <sub>2</sub>	media anual	40
	media de una hora	200
PM <sub>2,5</sub>	media anual	10
	media de 24 horas	25
PM <sub>10</sub>	media anual	20
	media de 24 horas	50
O <sub>3</sub>	media de ocho horas	100

FUENTE: WHO (2006)

Asimismo, el Decreto Supremo N° 003-2017-MINAM, Aprueban Estándares de Calidad Ambiental (ECA) para Aire y establecen Disposiciones Complementarias, establece los valores que no representen riesgo significativo para la salud de las personas ni al ambiente para los contaminantes SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>2,5</sub>, PM<sub>10</sub>, CO y O<sub>3</sub> (**Tabla 5**). Así mismo, en la **Tabla 6** se presenta las Guías de Calidad del Aire de la Organización Mundial de la Salud (OMS) para el SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>2,5</sub>, PM<sub>10</sub> y O<sub>3</sub>.

## 2.9 REGLAMENTO DE LOS NIVELES DE ESTADOS DE ALERTA POR CONTAMINACIÓN DEL AIRE

El Decreto Supremo N° 009-2003-SA, Reglamento de los Niveles de Estados de Alerta Nacionales para Contaminantes del Aire y su modificatoria el Decreto Supremo N° 012-2005-SA, establecen los niveles de alerta para el contaminante PM<sub>10</sub>. Este es un documento de gestión que tiene como fin la implementación de un conjunto de medidas predeterminadas para la prevención de riesgos a la salud y la exposición aguda de la población a los contaminantes atmosféricos. Los niveles de alerta para el PM<sub>10</sub> se presentan en la **Tabla 7**.

**Tabla 7: Niveles de Estados de Alerta Nacionales para Contaminantes del Aire**

Tipo de alerta	Material particulado (PM <sub>10</sub> )		Dióxido de azufre (SO <sub>2</sub> )		Monóxido de carbono (CO)	
Cuidado	>250 μ/m <sup>3</sup>	Promedio aritmético 24 horas	>500 μ/m <sup>3</sup>	Promedio móvil 3 horas	>15000 μ/m <sup>3</sup>	Promedio móvil 8 horas
Peligro	>350 μ/m <sup>3</sup>	Promedio aritmético 24 horas	>1500 μ/m <sup>3</sup>	Promedio móvil 3 horas	>20000 μ/m <sup>3</sup>	Promedio móvil 8 horas
Emergencia	>420 μ/m <sup>3</sup>	Promedio aritmético 24 horas	>2500 μ/m <sup>3</sup>	Promedio móvil 3 horas	>35000 μ/m <sup>3</sup>	Promedio móvil 8 horas

FUENTE: Decreto Supremo N° 009-2003-SA, Decreto Supremo N° 012-2005-SA

## 2.10 ÍNDICE DE CALIDAD DEL AIRE – INCA

A partir de la promulgación de la Resolución Ministerial N° 181-2016-MINAM, se establece el Índice de Calidad del Aire – INCA, que tiene como fin dar a conocer a la población el estado actual de la calidad del aire, presentando la información de forma clara y amigable, a ello se suma la constitución del Sistema de Información de Calidad del Aire – INFO AIRE PERÚ, que forma parte del SINIA, con el objetivo de fomentar el interés por la preservación

de un aire limpio, y sensibilizar a los tomadores de decisiones sobre la implementación de las acciones pertinentes que conlleven a la mejora en sus respectivas ciudades.

El INCA se divide en 4 categorías o calificaciones de la calidad del aire. La banda de color verde comprende un rango de valores del INCA de 0 a 50 y significa que la calidad del aire es buena; la banda de color amarillo comprende un rango de valores de 51 a 100 e indica una calidad moderada del aire; la banda de color anaranjado se encuentra comprendida entre los valores 101 y el valor umbral del estado de cuidado (VUEC) de cada contaminante, lo que nos indica que, la calidad del aire es mala; finalmente, el color rojo de la cuarta banda nos indica que, la calidad del aire es mayor al valor umbral del estado de cuidado del contaminante, a partir de este valor corresponde la aplicación de los Niveles de Estados de Alerta Nacionales por parte de la autoridad de Salud, Dirección General de Salud Ambiental en Inocuidad Alimentaria (DIGESA). La **Tabla 8** muestra los valores del INCA, con las 4 calificaciones y los colores utilizados para cada caso.

**Tabla 8: Valores del índice de calidad del aire**

Calificación	Valores del INCA	Colores
Buena	0-50	Verde
Moderada	51-100	Amarillo
Mala	101-VUEC*	Anaranjado
VUEC*	>VUEC*	Rojo

\*VUEC: Valor umbral del estado de cuidado

FUENTE: Resolución Ministerial N° 181-2016-MINAM

De acuerdo con la calificación del INCA la población sensible (niños, tercera edad, madres gestantes, personas con enfermedades respiratorias crónicas y cardiovasculares) y población en general deberán tomar en cuenta los cuidados y recomendaciones que se muestran en la **Tabla 9**.

En la **Tabla 10** se presenta los intervalos del INCA y los intervalos de las concentraciones de los contaminantes de estudio, con base a la Resolución Ministerial N° 181-2016-MINAM, el Decreto Supremo N° 003-2017-MINAM, el Decreto Supremo N° 009-2003-SA y el Decreto Supremo N° 012-2005-SA.

**Tabla 9: Cuidados y recomendaciones del índice de calidad del aire**

Calificación	Cuidados	Recomendaciones
Buena	La calidad del aire es satisfactoria y no representa un riesgo para la salud.	La calidad del aire es aceptable y cumple con el ECA de Aire. Puede realizar actividades al aire libre.
Moderada	La población sensible podría experimentar algunos problemas de salud.	La calidad del aire es aceptable y cumple con el ECA de Aire. Puede realizar actividades al aire libre con ciertas restricciones para la población sensible.
Mala	La población sensible podría experimentar problemas de salud. La población en general podría sentirse afectada.	Mantenerse atento a los informes de calidad del aire. Evitar realizar ejercicio y actividades al aire libre.
Umbral de cuidado	La concentración del contaminante puede causar efectos en la salud de cualquier persona y efectos serios en la población sensible.	Reportar a la Autoridad de Salud para que declare los Niveles de Estados de Alerta de acuerdo al Decreto Supremo N° 009-2003- SA y su modificatoria Decreto Supremo N° 012-2005-SA.

FUENTE: Resolución Ministerial N° 181-2016-MINAM

**Tabla 10: Índice de Calidad del Aire**

SO <sub>2</sub> promedio 24 horas		NO <sub>2</sub> promedio 1 hora		PM <sub>2,5</sub> promedio 24 horas		PM <sub>10</sub> promedio 24 horas		CO promedio 8 horas		O <sub>3</sub> promedio 8 horas	
I <sub>INCA</sub>	I <sub>c</sub>	I <sub>INCA</sub>	I <sub>c</sub>	I <sub>INCA</sub>	I <sub>c</sub>	I <sub>INCA</sub>	I <sub>c</sub>	I <sub>INCA</sub>	I <sub>c</sub>	I <sub>INCA</sub>	I <sub>c</sub>
0-50	0-125	0-50	0-100	0-50	0-25	0-50	0-50	0-50	0-5000	0-50	0-50
51-100	126-250	51-100	101-200	51-100	26-50	51-100	51-100	51-100	5001-10000	51-100	51-100
101-200	251-500	101-150	201-300	101-250	101-125	101-250	101-250	101-210	10001-15050	101-210	101-210
>200	>500	>150	>300	>250	>125	>250	>250	>210	>15050	>210	>210

I<sub>INCA</sub>: Intervalo del INCA; I<sub>c</sub>: Intervalo de concentraciones (µg/m<sup>3</sup>)

FUENTE: Elaborado con base a Resolución Ministerial N° 181-2016-MINAM, Decreto Supremo N° 003-2017-MINAM, Decreto Supremo N° 009-2003-SA, Decreto Supremo N° 012-2005-SA

## 2.11 ANÁLISIS ESTADÍSTICO MULTIVARIADO

El análisis multivariado es la rama del análisis estadístico que se centra en la investigación simultánea de dos o más características (variables) medidas en un conjunto de objetos (Kachigan 1991). De acuerdo con la definición, el elemento central es la relación simultánea entre las variables. Las técnicas multivariadas se diferencian de las univariadas y bivariadas en que dirigen su atención no en el análisis de la media y la varianza de una sola variable, o la correlación entre dos variables, sino al análisis de las covarianzas o correlaciones que presentan la relación entre tres o más variables (Aldás y Uriel 2017).



Las técnicas estadísticas multivariadas son un campo muy desarrollado y cuentan con muchos métodos diferentes. Muchos de estos son matemáticos. Afortunadamente, estos métodos se han programado para que puedan ejecutarse en un computador sin mucha dificultad (Zeltermán 2015).

## **2.12 TÉCNICAS ESTADÍSTICAS MULTIVARIADAS**

Se presentan, a continuación, las principales técnicas estadísticas multivariadas.

### **2.12.1 Análisis de componentes principales (ACP)**

El ACP es una técnica estadística que transforma un conjunto original de variables interrelacionadas en un nuevo conjunto, de un número igual de variables independientes no correlacionadas, denominadas componentes principales (CP). Estas CP son combinaciones lineales de las variables originales y se obtienen de tal manera que la primera CP explica la fracción más grande de la variabilidad de los datos originales. La segunda CP explica una fracción menor de la variación de datos que la primera CP y así sucesivamente. El número máximo de nuevas variables es equivalente al número de variables originales. La multicolinealidad que probablemente implica entre las variables originales puede eliminarse mediante la aplicación de la misma (Sousa *et al.* 2007).

El ACP es un método efectivo de clasificación, que generalmente se aplica un algoritmo de rotación, como la rotación *varimax* para minimizar la asociación entre las CP. De esta rotación se obtienen los factores de carga rotados que representan la contribución de cada variable a una CP específica (Sousa *et al.* 2007). Cuanto mayor es la carga factorial de esa variable, más es la contribución de esta, a la variación representada por la CP en particular (Jolliffe 2002).

Los valores de factores de carga que son superiores a 0.75 ( $> 0.75$ ) se consideran "fuertes", los valores que oscilan entre 0.50 y 0.75 ( $0.50 \geq \text{factor de carga} \geq 0.75$ ) se consideran como "moderados" y los valores que oscilan entre 0.30 y 0.49 ( $0.30 \geq \text{factor de carga} \geq 0.49$ ) se consideran como cargas de factor "débil" (Azid *et al.* 2015). En la práctica, solo se seleccionan cargas factoriales con valores absolutos superiores a 0.5 para la interpretación del componente principal (Azid *et al.* 2015).

Este método retiene la mayor parte de la información original y supera los problemas de correlación y superposición entre los datos originales. Los componentes principales (CP) se pueden expresar como en la Ecuación 1 a continuación:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \dots + a_{im}x_{mj} \quad (1)$$

donde,  $z$  es la puntuación del componente,  $a$  es la carga del componente,  $x$  es el valor medido de la variable,  $i$  es el número del componente,  $j$  es el número de muestra y  $m$  es el número total de variables.

### 2.12.2 Análisis de clúster jerárquico (ACJ)

El ACJ es un método de clasificación que se utiliza para dividir un conjunto de datos en clases o grupos (clúster). Su principal objetivo es establecer un conjunto de clúster de manera que los objetos de un mismo clúster (CL) sean similares entre sí y diferentes de los objetos ubicados en otro clúster, es decir, maximizar la varianza entre grupos y minimizar la varianza dentro del grupo (Pires *et al.* 2008; Lu *et al.* 2011). El número ideal de clúster se puede determinar gráficamente a través de un dendrograma, un diagrama de árbol comúnmente utilizado en el ACJ (Pires *et al.* 2008; Lu *et al.* 2011). Este método de clasificación también puede ser útil para la reducción de datos.

El ACJ es un método de identificación de patrones sin supervisión que divide un grupo grande en grupos más pequeños (Giordani *et al.* 2020). El ACJ permite identificar grupos invisibles que se ilustran con datos numéricos, simbólicos o estructurales, de modo que, los miembros de un grupo significativo comparten las similitudes entre ellos; mientras que, los grupos formados están bien diferenciados entre sí (Wehrens 2020). El ACJ se emplea en el conjunto de datos a través de diferentes métodos de enlace en base a distancias euclidianas, como una medida de la relación (Giordani *et al.* 2020).

El resultado de este método se representa por un método de árbol, que se conoce como un dendrograma. La distancia euclidiana (conocida como distancia de enlace) se contabiliza como  $D_{link} / D_{max}$ . Significa la medida entre las distancias de vinculación divididas por la distancia máxima. La medida se multiplicará por 100 como una forma de estandarizar la distancia de enlace representada sobre el eje y (Azid *et al.* 2015). La distancia euclidiana se puede definir por la Ecuación (2):

$$d(x, y) = \sum_{m=1}^p (x_m - y_m) \quad (2)$$

donde,  $d(x,y)$  es la distancia euclidiana entre dos elementos representados por  $x_m$  y  $y_m$ ;  $p$  es el espacio dimensional de las variables.

### 2.12.3 Regresión lineal múltiple (RLM)

La regresión lineal múltiple es una técnica estadística que nos permite predecir la variabilidad entre la variable independiente y la variable dependiente; así como calcular el porcentaje de la contribución de cada parámetro a la contaminación. El modelo general de regresión lineal múltiple tiene  $k$  variables independientes y  $n$  observaciones (Aldás y Uriel 2017).

El modelo generaliza la regresión lineal simple, en la cual cada valor de la variable independiente está asociado con un valor de la variable dependiente. El modelo de regresión se obtiene usando la Ecuación 3:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_{p-1} x_{p-1} + \varepsilon \quad (3)$$

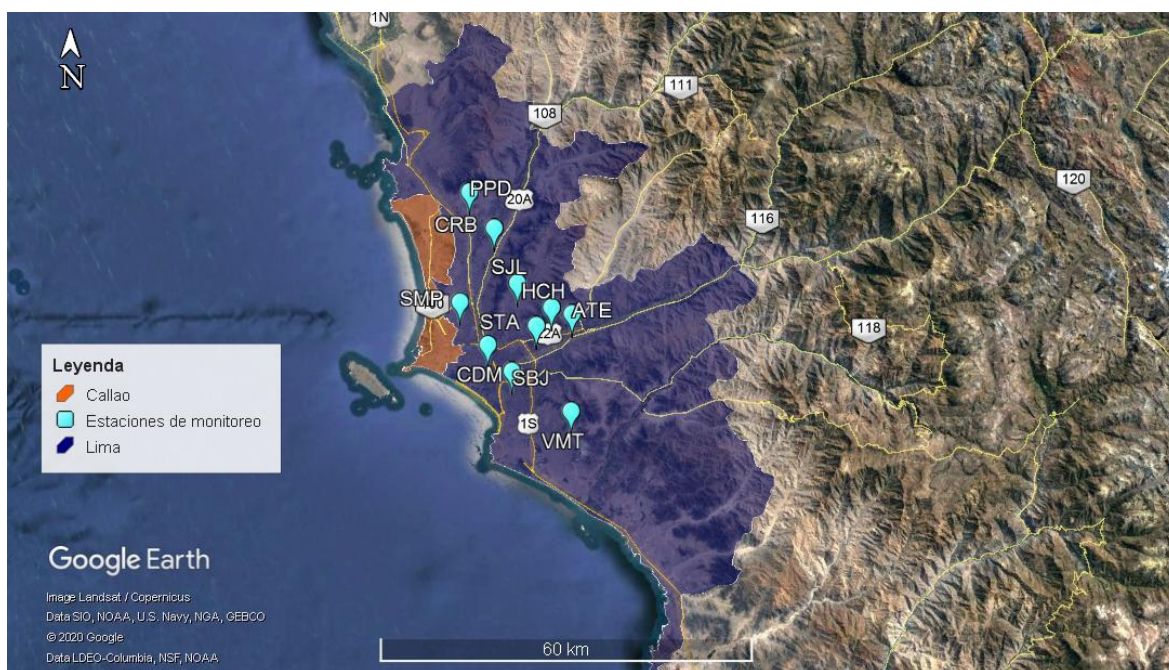
donde,  $Y$  es la variable de respuesta, y hay  $p - 1$  variables explicativas  $x_1, x_2, \dots, x_{p-1}$ , con  $p$  parámetros (coeficientes de regresión)  $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$  y  $\varepsilon$  es un error asociado con la regresión.

Para obtener la ecuación de regresión lineal más adecuada, el coeficiente de determinación ( $R^2$ ), el coeficiente de determinación ajustado ( $R^2$  ajustado) y la raíz del error cuadrático medio (RECM) son los indicadores que deben considerarse (Aldás y Uriel 2017). Desafortunadamente, el valor de  $R^2$  solo proporciona información sobre su buen funcionamiento en datos externos (Dominick *et al.* 2012). Sin embargo, los modelos que presentan los más altos valores de  $R^2$  (cerca de 1) son declarados como los mejores modelos lineales (Azid *et al.* 2015, Aldás y Uriel 2017).

### III. MATERIALES Y MÉTODOS

#### 3.1. DESCRIPCIÓN DEL ÁREA DE ESTUDIO

El presente estudio se realizó en el Área Metropolitana de Lima que comprende un total de 50 distritos, 43 distritos pertenecen a la Provincia de Lima y 7 distritos a la Provincia Constitucional del Callao. (Figura 2). Esta zona metropolitana se ubica geográficamente entre las Latitudes 11°45' y 12°24' Sur y las Longitudes 76°40' y 77°10' Oeste y para estudios atmosféricos se han zonificado tres cuencas atmosféricas, denominadas Cuenca del Río Chillón, Cuenca del Río Rímac y Cuenca del Río Lurín (DIGESA, 2012). Su geografía, en dirección noreste, abarca una gran llanura aluvial formada por los valles de los ríos antes mencionados, hasta alcanzar paisajes colinosos y finalmente montañosos en el extremo este, que pueden alcanzar alturas de más de 1000 metros (las primeras estribaciones de las laderas occidentales de los Andes centrales) (Silva *et al.* 2017).



**Figura 2: Área de estudio y localización de las estaciones de la Red de Monitoreo Automático de la Calidad del Aire del Área Metropolitana de Lima**

FUENTE: Silva *et al.* (2017), Google Earth (2020)

Se caracteriza por presentar diversos microclimas en su territorio, situación que es influenciada por la posición geográfica (latitud), la presencia de la cordillera de los Andes, su relieve, la corriente fría de Humboldt y el anticiclón del Pacífico Sur (APS), los cuales influyen en el comportamiento de las principales variables meteorológicas. Los vientos dominantes soplan desde el sur, suroeste y oeste alrededor de todo el año, y como resultado transportan los contaminantes emitidos hacia las zonas norte y este del AMLC (Sánchez *et al.* 2015, Silva *et al.* 2017, MINAM 2019).

### **3.2 UBICACIÓN DE LAS ESTACIONES DE MONITOREO Y DATOS DISPONIBLES**

La calidad del aire en Lima Metropolitana es monitoreada a través de dos instituciones. La primera, la Dirección General de Salud Ambiental e Inocuidad Alimentaria (DIGESA), adscrita al Ministerio de Salud, que se encarga de la vigilancia de la calidad del aire en Lima Metropolitana a través de siete estaciones de monitoreo, donde se utiliza comúnmente equipos activos y automáticos. La instalación de todas las estaciones se ha realizado en las azoteas del primer piso de los Centros de salud y hospitales de la red asistencial del Ministerio de Salud (DIGESA 2020).

La segunda institución es el Servicio Nacional de Meteorología e Hidrología del Perú - SENAMHI, adscrita al Ministerio del Ambiente (MINAM), que realiza la vigilancia de la meteorología y la calidad del aire en el AMLC a través de diez estaciones de monitoreo, conformando la Red de Monitoreo Automático de la Calidad del Aire (REMCA). Las estaciones de la REMCA operada por el SENAMHI se han instalado en edificios públicos, municipalidades y parques; y en todas las estaciones se utilizan equipos de monitoreo automáticos para gases y partículas (Sánchez-Ccoyllo *et al.* 2016, Silva *et al.* 2017).

Los sitios de la red se seleccionan con el objetivo de evaluar las áreas donde la población realiza sus actividades. Las estaciones cumplen con las regulaciones de salud y seguridad ocupacional bajo el marco de la normativa nacional. El sistema de energía eléctrica debe cumplir con los códigos eléctricos. En general, la temperatura del shelter debe mantenerse entre 20 y 30 °C. Las operaciones de las estaciones incluyen, visitas a la estación programadas regularmente, verificaciones de cero y span de los instrumentos, calibraciones, detección de fugas, mantenimiento preventivo y documentación. Las verificaciones se realizan con frecuencia semanal y/o mensual. Los registros se transmiten por telemetría a la sede del SENAMHI donde se validan, se corrigen entradas nulas, duplicados y/o anomalías

(Silva *et al.* 2017). La base de datos tiene una resolución de 1 hora. Las características generales de estas estaciones se muestran en la **Tabla 11**.

Los métodos de análisis de los contaminantes están de acuerdo con los Estándares de Calidad Ambiental (ECA) para Aire (MINAM 2017) y el Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire (MINAM 2019). Las concentraciones máscas de PM<sub>10</sub> se obtuvieron mediante las técnicas de Microbalanza oscilatoria y Atenuación de Rayos Beta, utilizando los equipos Monitor TEOM 1405 y Monitor 5014i de Thermo Scientific, respectivamente. Las concentraciones máscas de PM<sub>2,5</sub> se analizaron mediante la técnica Atenuación de Rayos Beta utilizando el equipo Monitor 5014i de Thermo Scientific. Las concentraciones máscas de NO<sub>2</sub> se obtuvieron mediante la técnica Fluorescencia UV utilizando el equipo Analizador 100E de Teledyne. Las concentraciones de O<sub>3</sub> se analizaron mediante la técnica de Fotometría UV utilizando el equipo Analizador 400E de Teledyne. Las concentraciones de CO se obtuvieron mediante la técnica Infrarrojo No Dispersivo utilizando el equipo Analizador T300 de Teledyne. Las concentraciones de SO<sub>2</sub> se determinaron mediante la técnica Fluorescencia UV utilizando el equipo Analizador 100E de Teledyne (Sánchez y Ordoñez 2016).

**Tabla 11: Características generales de la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de Lima**

Zona de Lima	N°	Estación	Distrito	Abreviatura	Norte	Este
Norte	1	Puente Piedra	Puente Piedra	PPD	8689197	274493
	2	Carabayllo	Carabayllo	CRB	8683451	278498
	3	San Martín de Porres	San Martín de Porres	SMP	8671605	273048
	4	Ate	Ate	ATE	8669832	291128
Este	5	Santa Anita	Santa Anita	STA	8667919	285386
	6	San Juan de Lurigancho	San Juan de Lurigancho	SJL	8674689	282309
	7	Huachipa	Huachipa	HCH	8670828	287828
Sur	8	Villa María del Triunfo	Villa María del Triunfo	VMT	8654310	291084
Centro	9	Campo de Marte	Jesús María	CDM	8664817	277593
	10	San Borja	San Borja	SBJ	8660605	281492

### 3.3. METODOLOGÍA

#### 3.3.1 Determinación de los patrones espaciales de los contaminantes atmosféricos (PM<sub>10</sub>, PM<sub>2,5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO y SO<sub>2</sub>) medidos en la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ)

Se procesaron los datos horarios proporcionados por el Servicio Nacional de Meteorología e Hidrología (SENAMHI) del Perú de la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del Área Metropolitana de Lima, de los parámetros PM<sub>2,5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO y variables meteorológicas durante el periodo 2015-2019, según estación de calidad de aire de procedencia y se integraron en una sola base de datos.

Se obtuvieron las estadísticas descriptivas de los datos de concentración de contaminantes atmosféricos. Se graficaron los diagramas de cajas y bigotes por contaminantes en las estaciones de monitoreo sin considerar los valores atípicos, los cuales son observaciones que se encuentran a por lo menos 1.5 veces del rango intercuartil (Q3 – Q1) del borde de la caja (Abdullah *et al.* 2019, Zamora *et al.* 2020).

Se empleó el método efectivo del ACP para transformar un grupo original de concentraciones de contaminantes potencialmente interrelacionadas en un nuevo grupo de variables independientes no correlacionadas, que se denominan componentes principales (CP). Se aplicó el algoritmo de rotación varimax para minimizar la asociación entre las CP (Sousa *et al.* 2007, Wang *et al.* 2018, Zhao *et al.* 2015). Las CP son combinaciones lineales de las concentraciones de contaminantes originales y están ordenadas de tal manera que la primera CP representa la mayor parte de la varianza en los datos originales, la segunda CP explica la segunda mayor proporción de la varianza, y así sucesivamente. Por lo tanto, la información principal de las concentraciones de contaminantes originales se conservó, al tiempo que se eliminó la colinealidad o la correlación entre las variables originales (Pires *et al.* 2008, Polanco 2016, Zhao *et al.* 2015). El método del ACP, se eligió para analizar los patrones espaciales de la Red de Monitoreo Automático de la Calidad del Aire (REMCA) en el Área Metropolitana de Lima en función de sus características de contaminación, proporcionando así, la base para la clasificación de las CP y los valores de factores de carga.

En el proceso del ACP, se eligió la función *na.omit* para manejar los valores faltantes y la función *varimax* para mitigar las asociaciones entre las variables. Se adoptó una

combinación de dos criterios para la extracción de las CP: primero se seleccionó las CP con un valor propio mayor o igual a 1 (criterio de Kaiser) (Pires *et al.* 2008, Polanco 2016, Zhao *et al.* 2015), luego se continuó eligiendo las CP hasta que explicaron más del 85 por ciento de la varianza acumulada. Todos los análisis se realizaron con el software estadístico R (R Core Team 2020), bajo su entorno de desarrollo integrado RStudio (RStudio 2020).

El ACJ es una técnica de clasificación que se utilizó para dividir el conjunto de concentraciones horarias de contaminantes en varios grupos (es decir, clúster). Se establecieron un conjunto de clases de modo que las observaciones en una clase dada sean similares entre sí, pero diferentes de las otras clases. Se utilizó un dendrograma para visualizar los resultados del ACJ (Lu *et al.* 2011, Zhao *et al.* 2015, Giordani *et al.* 2020).

Así también, los análisis se realizaron con el software estadístico R (R Core Team 2020), bajo el software de código abierto RStudio: Integrated Development Environment for R (RStudio 2020). En esta investigación, se empleó un algoritmo de aglomeración jerárquica para el agrupamiento, a través de la función *hclust*; y para el tratamiento de los datos faltantes se utilizó la función *na.omit*. El método de agrupamiento que se empleó es el enlace *ward* basado en una transformación estandarizada (puntajes Z). Se aplicó el enfoque de la distancia euclidiana para medir las distancias (es decir, las diferencias en términos de contaminación) entre las estaciones de monitoreo de calidad del aire. Las dos variables con la distancia euclidiana más pequeña, generalmente están emparentadas y constituyen un nuevo clúster. El procedimiento completo de agrupamiento que se utilizó en este estudio se presenta a continuación (Pires *et al.* 2008a, b):

- Paso 1: se calculó las distancias euclidianas entre todos los pares de estaciones;
- Paso 2: se agrupó las dos estaciones con la distancia euclidiana más baja en un nuevo clúster;
- Paso 3: se volvió a calcular las distancias euclidianas para las estaciones restantes y el clúster recién formado en el paso 2, luego se repitió el paso 2; y
- Paso 4: se repitió el paso 3 hasta que todas las estaciones evaluadas pertenezcan a un grupo.



### **3.3.2 Determinación de los patrones espaciales de las variables meteorológicas (temperatura de aire, humedad relativa, dirección y velocidad del viento y radiación solar) medidos en la red de estaciones meteorológicas automáticas del Área Metropolitana de Lima, mediante el análisis de componentes principales (ACP) y el análisis de clúster jerárquico (ACJ)**

Se verificó la idoneidad de las variables meteorológicas para el ACP, mediante las pruebas de Kaiser-Meyer-Olkin (KMO) y Bartlett. Para efectuar satisfactoriamente el análisis multivariado, la medida de KMO debe ser superior a 0.5 y la significancia de la prueba de Bartlett, debe ser menor a 0.05 (5%) (Ul-Saufie *et al.* 2013, Abdullah *et al.* 2018, Nazmi *et al.* 2020). Luego se estandarizaron en su media y desviación estándar por cada parámetro, obteniendo una media igual a cero y una desviación estándar igual a la unidad (Zhao *et al.* 2015). Después se empleó el ACP, para transformar las variables meteorológicas, que son variables potencialmente interrelacionadas, en un nuevo grupo de variables independientes no correlacionadas, que se denominan CP. Se aplicó el algoritmo de rotación varimax para minimizar la asociación entre las CP (Pires *et al.* 2009). Las CP son combinaciones lineales de las variables meteorológicas originales y están ordenadas de tal manera que la primera CP representa la mayor parte de la varianza en los datos originales, la segunda CP explica la segunda mayor proporción de la varianza, y así sucesivamente. Por lo tanto, la información primordial de las variables meteorológicas originales se conservó, al tiempo que se eliminó la colinealidad o la correlación entre las variables originales (Castro y Pires 2019). El método del ACP se utilizó para evaluar los patrones espaciales de la red de estaciones meteorológicas automáticas en el Área Metropolitana de Lima, proporcionando así la base para la clasificación de las CP y los valores de factores de carga.

En el desarrollo del ACP, se eligió la función *na.omit* para manejar los valores faltantes y la función *varimax* para mitigar las asociaciones entre las variables. Se estableció una combinación de dos criterios para la extracción de las CP: primero se seleccionó las CP con un valor propio mayor o igual a 1 (criterio de Kaiser) (Lu *et al.* 2011, Zhao *et al.* 2015), luego se continuó eligiendo las CP hasta que explicaron más del 85 por ciento de la varianza acumulada. Todos los análisis se realizaron con el software estadístico R (R Core Team 2020), bajo el software de código abierto RStudio: Integrated Development Environment for R (RStudio, 2020).

Se utilizó el ACJ para dividir el conjunto de variables meteorológicas en varios grupos (es decir, clúster). Se estableció un conjunto de clases de modo que las observaciones en una clase dada sean similares entre sí, pero diferentes de las otras clases. Se presenta un dendrograma para representar los resultados del ACJ (Zhao *et al.* 2015, Giordani *et al.* 2020).

De igual manera, todos los análisis se realizaron con el software estadístico R (R Core Team 2020) y su entorno de desarrollo integrado Rstudio (RStudio 2020). En esta investigación, se empleó un algoritmo de aglomeración jerárquica para el agrupamiento, a través de la función *hclust*; y para el tratamiento de los datos faltantes se utilizó la función *na.omit*. El método de agrupamiento que se empleó es el enlace ward basado en una transformación estandarizada (puntajes Z). Se empleó el enfoque de la distancia euclidiana para medir las diferencias o similitudes entre las estaciones meteorológicas automáticas. Las dos variables con la distancia euclidiana más baja generalmente están vinculadas y forman un nuevo clúster. El procedimiento completo de agrupamiento que se utilizó en este estudio se presenta a continuación (Pires *et al.* 2008a, b):

- Paso 1: se calculó las distancias euclidianas entre todos los pares de estaciones;
- Paso 2: se agrupó las dos estaciones con la distancia euclidiana más baja en un nuevo clúster;
- Paso 3: se volvió a calcular las distancias euclidianas para las estaciones restantes y el clúster recién formado en el paso 2, y luego se repitió el paso 2; y
- Paso 4: se repitió el paso 3 hasta que todas las estaciones evaluadas pertenezcan a un grupo.

### **3.3.3 Pronóstico de las concentraciones de material particulado (PM<sub>10</sub> y PM<sub>2,5</sub>) y gases (NO<sub>2</sub>, O<sub>3</sub>, CO y SO<sub>2</sub>) en el área de Lima Metropolitana a través de la regresión lineal múltiple (RLM)**

La regresión lineal múltiple (RLM) se utilizó para investigar la relación entre los parámetros de contaminación del aire (variables dependientes) y las variables meteorológicas (variable independiente) ajustando una ecuación lineal por mínimos cuadrados a los datos horarios observados (Azid *et al.* 2015, Wehrens 2020) proporcionando el porcentaje de la contribución de cada parámetro a la contaminación atmosférica (Azid *et al.* 2015). Se comparó el modelo realizado con los parámetros originales de los contaminantes atmosféricos y variables meteorológicas con el modelo elaborado en función de los

parámetros más significativos de los contaminantes atmosféricos y variables meteorológicas, con el objetivo de conseguir un mejor modelo dentro de los grupos.

Con la finalidad de cumplir el tercer objetivo de la presente investigación, se trabajó con la estación ATE, la cual es una de las estaciones de monitoreo que presenta los mayores niveles contaminantes de PM<sub>10</sub>, PM<sub>2,5</sub>, NO<sub>2</sub>, CO y SO<sub>2</sub> (Silva *et al.* 2017, 2018). Los pronósticos generados a una escala temporal horaria se obtuvieron en base a dos tipos de modelos (modelo I y modelo II) implementados para cada uno de los contaminantes. El primer modelo (modelo I) se basó en la predicción del contaminante tomando como predictores a las concentraciones del mismo contaminante de las otras nueve estaciones de monitoreo. El esquema del modelo I se presenta en la **Tabla 12**. Mientras que, el segundo modelo (modelo II) se implementó tomando como variables predictores a las variables meteorológicas y contaminantes de la misma estación de calidad de aire. El diseño del modelo II se presenta en la **Tabla 13**.

**Tabla 12: Esquema del primer modelo empleado para el pronóstico de las concentraciones horarias de los contaminantes atmosféricos**

Variable respuesta		Variables predictores								
Estación	ATE	STA	SMP	SJL	SBJ	PPD	HCH	VMT	CRB	CDM
	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>	PM <sub>10</sub>
Contaminante	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>	PM <sub>2,5</sub>
	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>	NO <sub>2</sub>
	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>	O <sub>3</sub>
	CO	CO	CO	CO	CO	CO	CO	CO	CO	CO
	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>	SO <sub>2</sub>

Todas las concentraciones en unidades de µg/m<sup>3</sup>

Para la implementación de todos los modelos generados, se procedió a dividir el conjunto de datos en dos grupos, uno de entrenamiento (80%) y el otro de validación (20%). En todos los modelos desarrollados se consideraron únicamente como predictores aquellos que eran estadísticamente significativos al 5%, mediante la aplicación de la prueba de t-student y el análisis de varianza (ANOVA) (Abdullah *et al.* 2019; Al-Karkhi y Alqaraghuli, 2019). Esto ocasionó que los modelos sean también estadísticamente significativos y válidos con un nivel de confianza de 95% a más. La validación consistió en aplicar, el modelo entrenado para el pronóstico, con los predictores no utilizados en la etapa de entrenamiento cuyos

resultados nos dan los pronósticos del contaminante, los cuales fueron utilizados para comparar con los observados mediante una regresión lineal simple. Previo a la partición del conjunto de datos, se trataron los valores atípicos, reemplazando estos de la siguiente manera, los valores atípicos superiores fueron sustituidos por el percentil 95; mientras que, los valores atípicos inferiores se reemplazaron por el percentil 5.

**Tabla 13: Esquema del segundo modelo empleado para el pronóstico de las concentraciones horarias de los contaminantes atmosféricos**

Variable respuesta		Variables predictores									
Estación	ATE	ATE	ATE	ATE	ATE	ATE	ATE	ATE	ATE	ATE	ATE
	PM <sub>10</sub>	PM <sub>2,5</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	SO <sub>2</sub>	T	HR	DV	VV	RS
	PM <sub>2,5</sub>	PM <sub>10</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	SO <sub>2</sub>	T	HR	DV	VV	RS
Contaminante	NO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2,5</sub>	O <sub>3</sub>	CO	SO <sub>2</sub>	T	HR	DV	VV	RS
	O <sub>3</sub>	PM <sub>10</sub>	PM <sub>2,5</sub>	NO <sub>2</sub>	CO	SO <sub>2</sub>	T	HR	DV	VV	RS
	CO	PM <sub>10</sub>	PM <sub>2,5</sub>	NO <sub>2</sub>	O <sub>3</sub>	SO <sub>2</sub>	T	HR	DV	VV	RS
	SO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2,5</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	T	HR	DV	VV	RS

T: temperatura del aire (°C); HR: humedad relativa (%); DV: dirección del viento (°); VV: velocidad del viento (m/s); RS: radiación solar (Wh/m<sup>2</sup>)

El valor de  $r^2$  ajustado se calculó considerando todas las posibles variables, ya que el  $r^2$  tiende a sobreestimar el éxito del modelo cuando se aplica al mundo real. Mientras tanto, se utilizó el error cuadrático medio (ECM), y la raíz del error cuadrático medio (RECM); para medir el error residual, ambos se tomaron en cuenta para la estimación de la diferencia media entre el valor observado y el valor pronosticado de las concentraciones de material particulado y gases. Se consideró como el mejor modelo, aquel que obtuvo la RECM más pequeño y el valor de  $r^2$  ajustado más cercano a 1, durante la etapa de validación (Azid *et al.* 2015, Karl y Hlávka 2015).

Una vez que se encontró el modelo adecuado éste se utilizó para pronosticar las concentraciones horarias de material particulado y gases en Lima Metropolitana, los cuales serán los insumos para los tomadores de decisiones en la prevención y control de la contaminación del aire en la ciudad.

## IV. RESULTADOS Y DISCUSIÓN

### 4.1 DETERMINACIÓN DE LOS PATRONES ESPACIALES DE LOS CONTAMINANTES ATMOSFÉRICOS MEDIDOS EN LA REMCA DEL ÁREA METROPOLITANA DE LIMA MEDIANTE EL ACP Y EL ACJ

En la **Tabla 14** se presentan las estadísticas descriptivas principales para las concentraciones de los contaminantes del presente estudio. Las mayores concentraciones en promedio se encuentran en la estación ATE; mientras que, las menores concentraciones en la estación CDM. Estos niveles están de acuerdo con las fuentes de emisión, condiciones meteorológicas y geográficas específicas de cada zona del AMLC (Pacsi 2016, Silva *et al.* 2017).

Las concentraciones promedio de PM<sub>10</sub> son más altas en las estaciones de monitoreo VMT (126.7 µg/m<sup>3</sup>), ATE (113.3 µg/m<sup>3</sup>), PPD (108.9 µg/m<sup>3</sup>) y HCH (96 µg/m<sup>3</sup>). Las concentraciones intermedias, las presentan las estaciones SJL (84.0 µg/m<sup>3</sup>), CRB (83.7 µg/m<sup>3</sup>) y STA (74.4 µg/m<sup>3</sup>). Y las concentraciones bajas de PM<sub>10</sub>, se presentan en las estaciones SBJ (53.0 µg/m<sup>3</sup>), SMP (46.1 µg/m<sup>3</sup>) y CDM (34.0 µg/m<sup>3</sup>). Similares resultados se encontraron en estudios anteriores (Silva *et al.* 2017, Espinoza 2018).

Para el PM<sub>2,5</sub>, generalmente se observa un patrón similar, aunque las concentraciones promedio entre las estaciones son más similares y están en el rango de 16.8 a 41.2 µg/m<sup>3</sup>. Las concentraciones promedias de mayor a menor de PM<sub>2,5</sub> se observan en ATE (41.2 µg/m<sup>3</sup>), SJL (31.8 µg/m<sup>3</sup>), PPD (31.7 µg/m<sup>3</sup>), STA (30.4 µg/m<sup>3</sup>), HCH (29.9 µg/m<sup>3</sup>), CRB (26.3 µg/m<sup>3</sup>), VMT (25.4 µg/m<sup>3</sup>), SMP (17.3 µg/m<sup>3</sup>), SBJ (16.8 µg/m<sup>3</sup>) y CDM (16.8 µg/m<sup>3</sup>). Se demarcan tres grupos de estaciones en términos de la magnitud de la concentración de PM<sub>2,5</sub>. El primer grupo, formado por las estaciones ATE, SJL, PPD y STA, presenta las mayores concentraciones; el segundo grupo, formado por las estaciones HCH, CRB y VMT, presenta concentraciones intermedias; y el tercer grupo, formado por las estaciones SMP, SBJ y CDM, tiene concentraciones más bajas. La principal fuente de emisión de PM<sub>2,5</sub> en el AMLC es el parque automotor, que contribuye con el 57.4% del total de emisiones para este

contaminante; seguido de las fuentes puntuales (26.3%) y finalmente las fuentes de área (16.3%) (MINAM 2019).

En el caso de NO<sub>2</sub>, las concentraciones promedio son más altas en las estaciones ATE (43.5 µg/m<sup>3</sup>), HCH (31.8 µg/m<sup>3</sup>) y SJL (31.7 µg/m<sup>3</sup>). Les siguen con concentraciones intermedias, las estaciones PPD (29.1 µg/m<sup>3</sup>), SMP (24.1 µg/m<sup>3</sup>), SBJ (24.0 µg/m<sup>3</sup>) y STA (23.2 µg/m<sup>3</sup>). Mientras que, las concentraciones multianuales de NO<sub>2</sub> más bajas, se presentan en las estaciones CDM (21.2 µg/m<sup>3</sup>), CRB (21.0 µg/m<sup>3</sup>) y VMT (17.5 µg/m<sup>3</sup>). La principal fuente de emisión de NO<sub>2</sub> en el AMLC es el parque automotor, que contribuye con el 98.6% del total de emisiones para este contaminante (MINAM 2019).

Las concentraciones promedio de O<sub>3</sub> para las estaciones de monitoreo de mayor a menor son CDM (20.9 µg/m<sup>3</sup>), SMP (17.7 µg/m<sup>3</sup>), HCH (17.1 µg/m<sup>3</sup>), CRB (17.0 µg/m<sup>3</sup>), SBJ (14.6 µg/m<sup>3</sup>), VMT (14.6 µg/m<sup>3</sup>), PPD (13.1 µg/m<sup>3</sup>), SJL (12.4 µg/m<sup>3</sup>), STA (11.8 µg/m<sup>3</sup>) y ATE (11.2 µg/m<sup>3</sup>). Se observan tres grupos de estaciones en términos de la magnitud de la concentración de O<sub>3</sub>. El primer grupo, formado por la estación CDM, presenta la mayor concentración; el segundo grupo, formado por las estaciones SMP, HCH y CRB, presenta concentraciones intermedias; y el tercer grupo, formado por las estaciones SBJ, VMT, PPD, SJL, STA y ATE, tiene concentraciones más bajas.

Para el CO, los mayores valores de concentración se encuentran en las estaciones ATE (1490.9 µg/m<sup>3</sup>) y PPD (1017.1 µg/m<sup>3</sup>); le siguen con concentraciones intermedias las estaciones SJL (902.9 µg/m<sup>3</sup>), HCH (886.9 µg/m<sup>3</sup>), STA (881.6 µg/m<sup>3</sup>), SBJ (807.8 µg/m<sup>3</sup>), VMT (788.9 µg/m<sup>3</sup>) y CRB (771.7 µg/m<sup>3</sup>); y finalmente, con menores concentraciones las estaciones SMP (581.8 µg/m<sup>3</sup>) y CDM (579.3 µg/m<sup>3</sup>). La principal fuente de emisión de CO en el AMLC es el parque automotor, que contribuye con el 96.5% del total de emisiones para este contaminante (MINAM 2019).

Las concentraciones promedio de SO<sub>2</sub> para las estaciones de monitoreo son más altas para ATE (21.8 µg/m<sup>3</sup>) y HCH (20.0 µg/m<sup>3</sup>); mientras que, concentraciones intermedias se encuentran en las estaciones PPD (12.1 µg/m<sup>3</sup>), SJL (11.0 µg/m<sup>3</sup>), STA (10.5 µg/m<sup>3</sup>) y SBJ (9.9 µg/m<sup>3</sup>); finalmente, las menores concentraciones se encuentran en las estaciones SMP (8.6 µg/m<sup>3</sup>), VMT (7.8 µg/m<sup>3</sup>), CRB (6.5 µg/m<sup>3</sup>) y CDM (5.5 µg/m<sup>3</sup>). La principal fuente de emisión de SO<sub>2</sub> en el AMLC son el parque automotor, que contribuye con el 96.5% del total de emisiones para este contaminante; seguido de las fuentes de área (3.4%) y finalmente las fuentes puntuales (0.2%) (MINAM 2019).

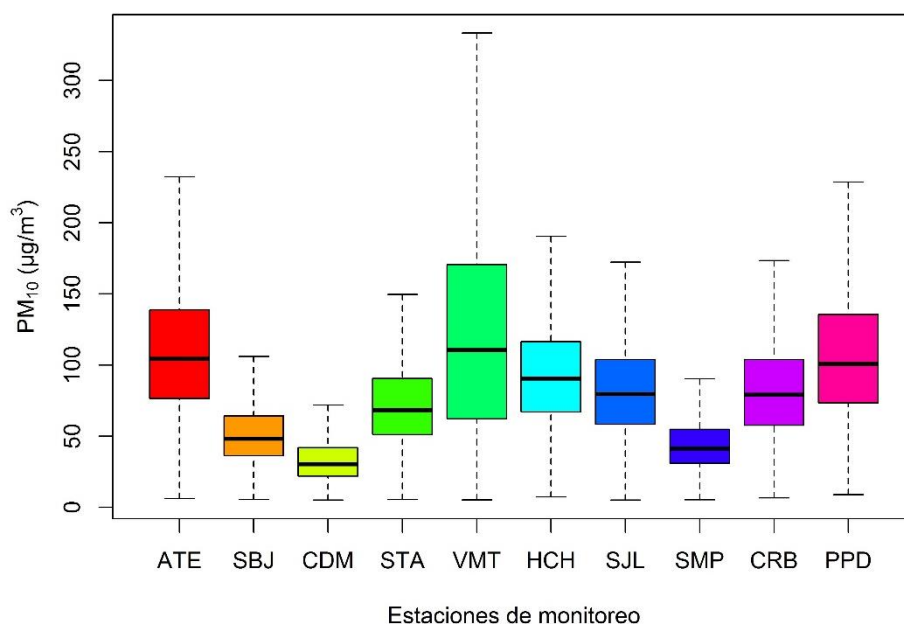
**Tabla 14: Estadísticas descriptivas de las concentraciones de los contaminantes atmosféricos**

Estación	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD	
PM <sub>10</sub>	Min	6.1	5.5	5.1	5.5	5.2	7.3	5.1	5.4	6.6	8.9
	Media	113.3±0.3	52.9±0.2	34.0±0.1	74.4±0.2	126.7±0.5	96.0±0.3	84.0±0.2	46.1±0.1	83.7±0.2	108.8±0.3
	Max	931.0	463.6	259.0	715.0	974.0	621.0	564.4	488.0	712.7	832.0
	DE	56.3	25.7	17.6	35.5	87.8	45.9	37.5	24.6	37.7	51.2
PM <sub>2,5</sub>	Min	3.0	3.0	3.0	3.1	3.0	3.0	3.0	3.0	3.0	3.0
	Media	41.2±0.2	16.8±0.1	16.8±0.1	30.4±0.1	25.4±0.1	29.9±0.2	31.8±0.1	17.3±0.1	26.3±0.1	31.7±0.1
	Max	786.9	223.8	98.7	535.2	191.9	336.2	720.7	423.4	285.6	455.6
	DE	25.4	9.8	9.6	17.0	15.1	21.8	18.7	10.4	17.2	15.0
NO <sub>2</sub>	Min	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.1
	Media	43.4±0.2	24.0±0.1	21.2±0.1	23.2±0.1	17.5±0.1	31.8±0.1	31.7±0.1	24.1±0.1	20.9±0.1	29.1±0.1
	Max	197.2	124.7	162.1	231.2	103.0	173.5	170.1	172.8	145.2	161.7
	DE	24.5	15.2	17.9	18.5	12.7	16.6	16.7	17.7	13.2	15.2
O <sub>3</sub>	Min	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	Media	11.2±0.1	14.6±0.1	20.9±0.1	11.8±0.1	14.6±0.1	17.1±0.1	12.4±0.1	17.7±0.1	17.0±0.1	13.1±0.1
	Max	136.0	108.7	117.1	152.6	113.9	165.8	118.1	125.3	150.3	148.8
	DE	12.7	10.5	18.1	14.2	12.3	19.9	16.0	13.7	20.1	16.1
CO	Min	34.5	0.7	0.1	0.3	0.2	0.1	1.3	0.7	5.6	0.5
	Media	1490.9±4.2	807.8±3.2	579.3±3.8	881.6±2.6	788.9±2.0	886.9±3.6	902.9±4.3	581.8±1.7	771.7±1.9	1017.1±3.0
	Max	5917.3	3671.9	4743.8	4202.1	3251.0	4547.1	5225.6	3122.7	4186.0	4534.4
	DE	723.3	565.5	644.2	451.6	328.4	616.4	741.2	319.0	342.9	503.4
SO <sub>2</sub>	Min	0.3	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.1
	Media	21.8±0.2	9.9±0.1	5.5±0.0	10.5±0.1	7.8±0.1	19.9±0.1	10.9±0.0	8.6±0.1	6.5±0.0	12.1±0.1
	Max	196.7	65.0	38.1	174.8	257.8	196.2	101.9	130.3	61.0	186.0
	DE	18.8	8.7	3.7	10.7	10.7	14.8	6.4	8.9	3.9	13.3

DE: desviación estándar

Es de resaltar que, las estaciones con mayor concentración de contaminantes atmosféricos se ubican en la zona este de la ciudad, es decir, la parte más alta; mientras que, las de menor concentración se encuentran en la parte occidental, que es la parte más baja de la ciudad (**Figura 2**); esta tendencia se debe al patrón de vientos locales persistentes que ingresan desde la costa con dirección sur-suroeste, provocando que las cargas contaminantes se transfieran a las zonas este y noreste, que son áreas críticas de deposición (Silva *et al.* 2017, Espinoza 2018).

En las **Figuras 3 al 8** se presentan los diagramas de cajas y bigotes de las concentraciones de contaminantes atmosféricos para las diez estaciones de monitoreo de calidad del aire administradas por el SENAMHI, sin considerar los valores atípicos, los cuales son observaciones que se encuentran a por lo menos 1.5 veces el rango intercuartil ( $Q3 - Q1$ ) del borde de la caja (Abdullah *et al.* 2019).

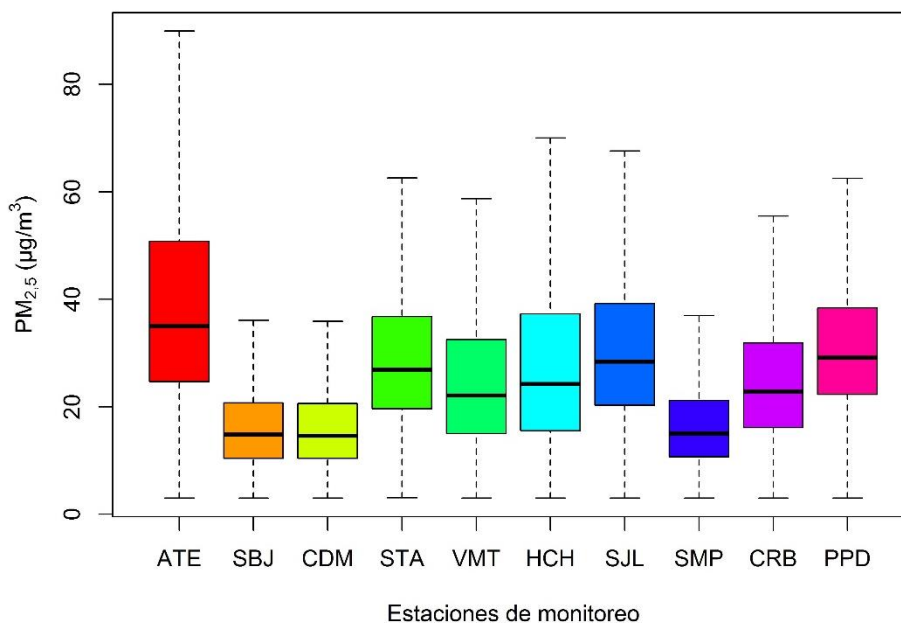


**Figura 3: Diagrama de cajas y bigotes para las concentraciones de PM<sub>10</sub>**

A continuación, se presentan el primer, segundo y tercer cuartil, respectivamente, de las concentraciones de PM<sub>10</sub> en las siguientes estaciones: VMT (62.1, 110.5 y 170.7 µg/m<sup>3</sup>), ATE (76.5, 104.4 y 138.8 µg/m<sup>3</sup>), PPD (73.4, 100.7 y 135.5 µg/m<sup>3</sup>), HCH (67.0, 90.5 y 116.4 µg/m<sup>3</sup>), SJL (58.3, 79.6 y 103.9 µg/m<sup>3</sup>), CRB (57.6, 79.0 y 103.9 µg/m<sup>3</sup>), STA (51.1, 68.4 y 90.5 µg/m<sup>3</sup>), SBJ (36.3, 48.2 y 64.2 µg/m<sup>3</sup>), SMP (30.9, 41.2 y 54.7 µg/m<sup>3</sup>) y CDM (21.9, 30.3 y 41.9 µg/m<sup>3</sup>) tal como se observa en la **Figura 3**. Donde, por ejemplo, para la estación ATE, el primer cuartil es 62.1 µg/m<sup>3</sup>, lo que significa que el 25% de los datos de



concentraciones de  $PM_{10}$  son igual o menor a  $62.1 \mu\text{g}/\text{m}^3$ . El segundo cuartil es  $110.5 \mu\text{g}/\text{m}^3$ , entonces el 50% de los datos es igual o menor a  $110.5 \mu\text{g}/\text{m}^3$  y; por último, el tercer cuartil es  $170.7 \mu\text{g}/\text{m}^3$ , se puede concluir que el 75 % de los datos son iguales o menores que  $170.7 \mu\text{g}/\text{m}^3$ . Y así sucesivamente para el resto de estaciones de monitoreo.



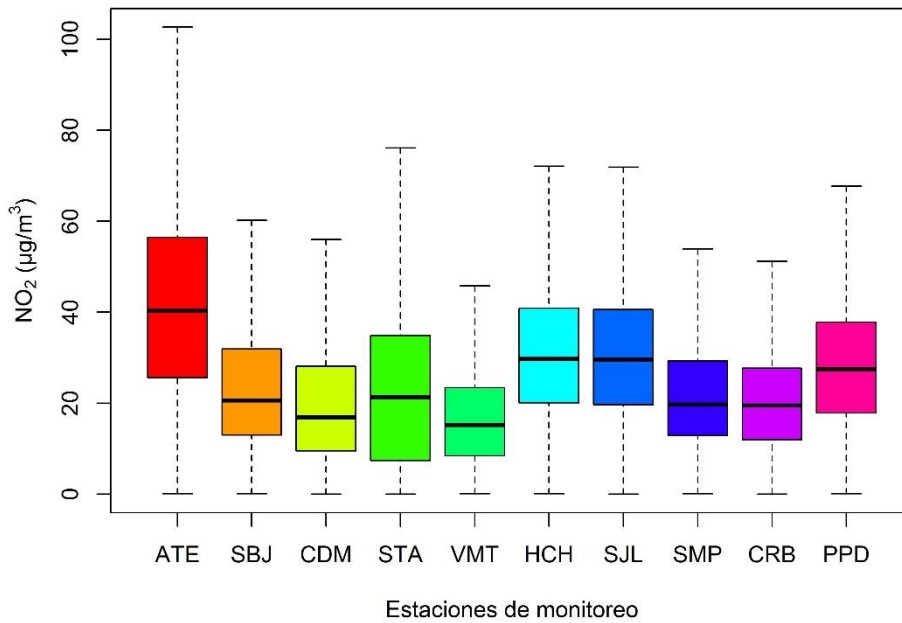
**Figura 4: Diagrama de cajas y bigotes para las concentraciones de  $PM_{2.5}$**

Para las concentraciones del  $PM_{2.5}$  se muestran, el primero, segundo y tercer cuartil, respectivamente, de las estaciones ATE ( $24.7, 35.0$  y  $50.8 \mu\text{g}/\text{m}^3$ ), SJL ( $20.3, 28.4$  y  $39.2 \mu\text{g}/\text{m}^3$ ), PPD ( $22.3, 29.1$  y  $38.4 \mu\text{g}/\text{m}^3$ ), STA ( $19.6, 26.9$  y  $36.8 \mu\text{g}/\text{m}^3$ ), HCH ( $15.5, 24.2$  y  $37.3 \mu\text{g}/\text{m}^3$ ), CRB ( $16.1, 22.8$  y  $31.9 \mu\text{g}/\text{m}^3$ ), VMT ( $15.0, 22.1$  y  $32.5 \mu\text{g}/\text{m}^3$ ), SMP ( $10.7, 15.0$  y  $21.2 \mu\text{g}/\text{m}^3$ ), SBJ ( $10.4, 14.8$  y  $20.7 \mu\text{g}/\text{m}^3$ ); y CDM ( $10.4, 14.6$  y  $20.6 \mu\text{g}/\text{m}^3$ ) tal como se observa en la **Figura 4**.

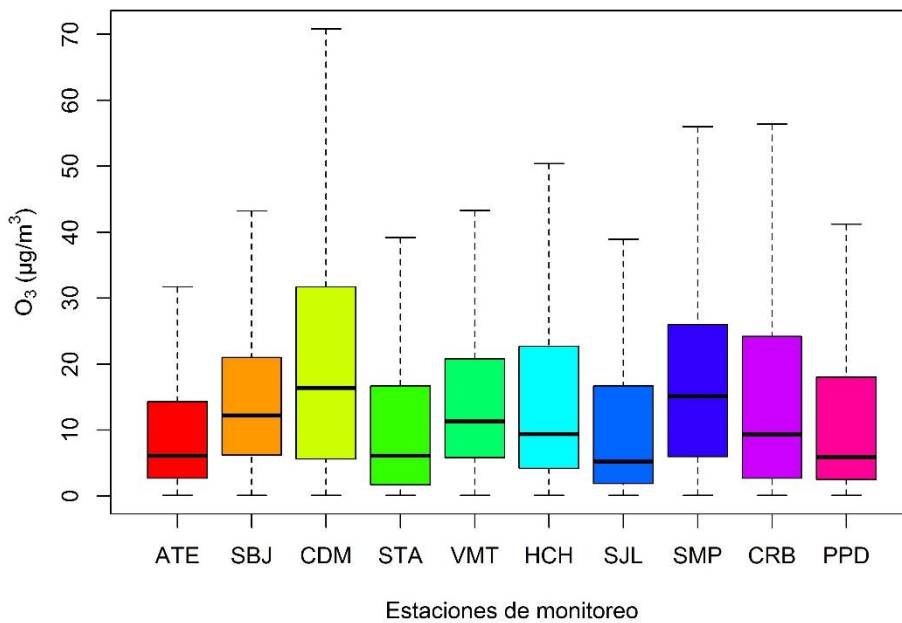
Para el  $NO_2$  se presenta, el primero, segundo y tercer cuartil, respectivamente, de las estaciones ATE ( $25.6, 40.3$  y  $56.5 \mu\text{g}/\text{m}^3$ ), HCH ( $20.1, 29.7$  y  $40.9 \mu\text{g}/\text{m}^3$ ), SJL ( $19.7, 29.6$  y  $40.6 \mu\text{g}/\text{m}^3$ ), PPD ( $17.8, 27.5$  y  $37.8 \mu\text{g}/\text{m}^3$ ), SMP ( $12.9, 19.7$  y  $29.3 \mu\text{g}/\text{m}^3$ ), SBJ ( $13.0, 20.6$  y  $31.9 \mu\text{g}/\text{m}^3$ ), STA ( $7.4, 21.3$  y  $34.9 \mu\text{g}/\text{m}^3$ ), CDM ( $9.5, 16.9$  y  $28.1 \mu\text{g}/\text{m}^3$ ), CRB ( $12.0, 19.5$  y  $27.7 \mu\text{g}/\text{m}^3$ ) y VMT ( $8.4, 15.2$  y  $23.4 \mu\text{g}/\text{m}^3$ ) tal como se muestra en la **Figura 5**.

Para el  $O_3$  se muestra, el primero, segundo y tercer cuartil, respectivamente, de las estaciones CDM ( $5.6, 16.4$  y  $31.7 \mu\text{g}/\text{m}^3$ ), SMP ( $6.0, 15.1$  y  $26.0 \mu\text{g}/\text{m}^3$ ), HCH ( $4.2, 9.4$  y  $22.7 \mu\text{g}/\text{m}^3$ ), CRB ( $2.7, 9.3$  y  $24.2 \mu\text{g}/\text{m}^3$ ), SBJ ( $6.2, 12.2$  y  $21.0 \mu\text{g}/\text{m}^3$ ), VMT ( $5.8, 11.3$  y  $20.8 \mu\text{g}/\text{m}^3$ ),

PPD (2.5, 5.9 y 18.0  $\mu\text{g}/\text{m}^3$ ), SJL (1.9, 5.2 y 16.7  $\mu\text{g}/\text{m}^3$ ), STA (1.7, 6.1 y 16.7  $\mu\text{g}/\text{m}^3$ ); y ATE (2.7, 6.1 y 14.3  $\mu\text{g}/\text{m}^3$ ) tal como se observa en la **Figura 6**.



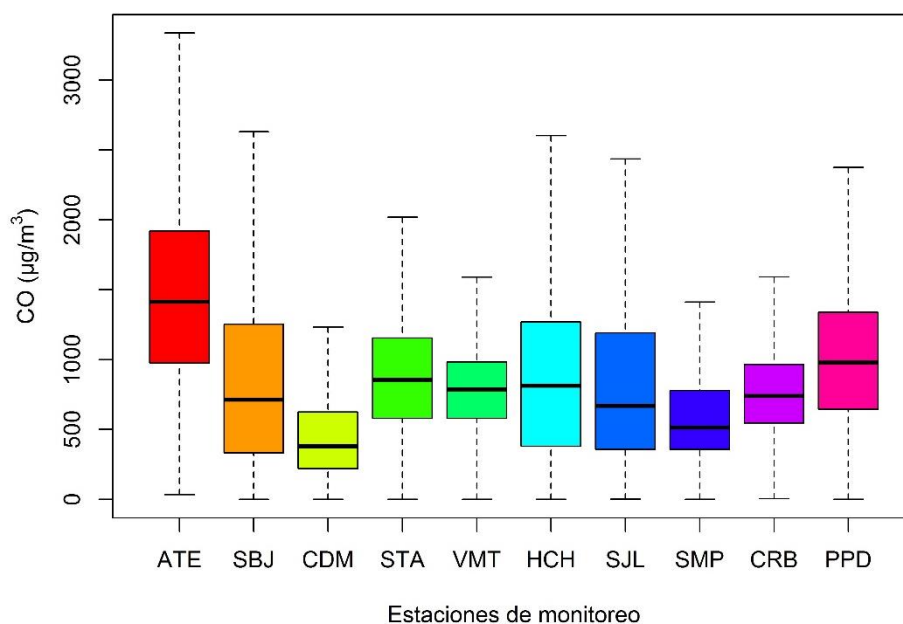
**Figura 5: Diagrama de cajas y bigotes para las concentraciones de NO<sub>2</sub>**



**Figura 6: Diagrama de cajas y bigotes para las concentraciones de O<sub>3</sub>**

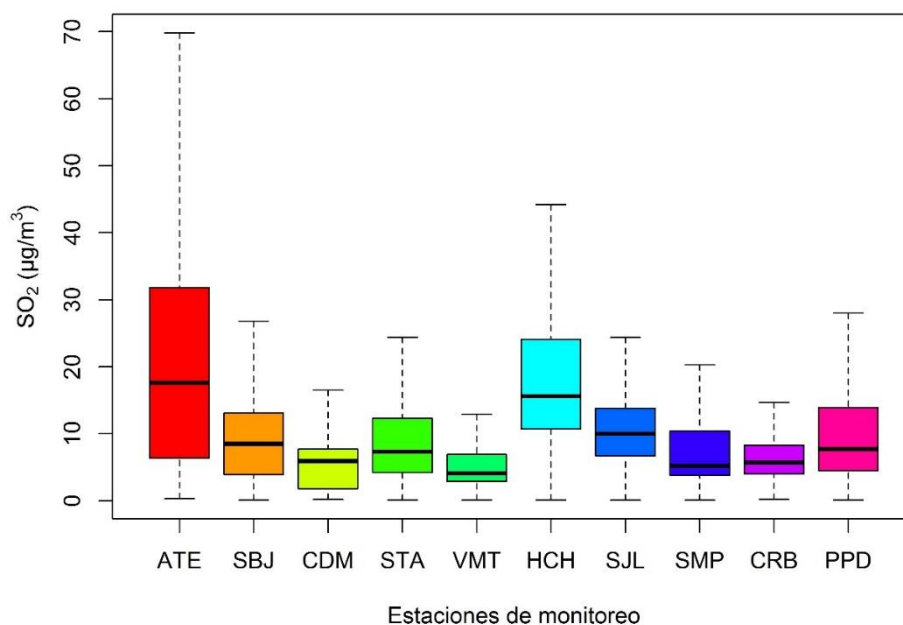
Para el CO se presenta, el primero, segundo y tercer cuartil, respectivamente, de las estaciones ATE (975.2, 1413.3 y 1920.5  $\mu\text{g}/\text{m}^3$ ), PPD (644.4, 978.6 y 1338.6  $\mu\text{g}/\text{m}^3$ ), SJL (358.8, 669.3 y 1190.2  $\mu\text{g}/\text{m}^3$ ), HCH (379.5, 813.0 y 1269.6  $\mu\text{g}/\text{m}^3$ ), STA (579.6, 855.6 y 1155.7  $\mu\text{g}/\text{m}^3$ ), SBJ (333.5, 713.0 y 1253.5  $\mu\text{g}/\text{m}^3$ ), VMT (578.7, 786.6 y 983.6  $\mu\text{g}/\text{m}^3$ ), CRB

(546.2, 739.4 y 964.8  $\mu\text{g}/\text{m}^3$ ), SMP (357.6, 513.0 y 779.2  $\mu\text{g}/\text{m}^3$ ); y CDM (218.5, 379.5 y 625.3  $\mu\text{g}/\text{m}^3$ ) tal como se muestra en la **Figura 7**.



**Figura 7: Diagrama de cajas y bigotes para las concentraciones de CO**

Para el  $\text{SO}_2$  se muestra, el primero, segundo y tercer cuartil, respectivamente, de las estaciones ATE (6.4, 17.6 y 31.8  $\mu\text{g}/\text{m}^3$ ), HCH (10.7, 15.6 y 24.1  $\mu\text{g}/\text{m}^3$ ), PPD (4.5, 7.7 y 13.9  $\mu\text{g}/\text{m}^3$ ), SJL (6.7, 10.0 y 13.8  $\mu\text{g}/\text{m}^3$ ), STA (4.2, 7.3 y 12.3  $\mu\text{g}/\text{m}^3$ ), SBJ (3.9, 8.5 y 13.1  $\mu\text{g}/\text{m}^3$ ), SMP (3.8, 5.2 y 10.4  $\mu\text{g}/\text{m}^3$ ), VMT (2.9, 4.1 y 6.9  $\mu\text{g}/\text{m}^3$ ), CRB (4.0, 5.7 y 8.3  $\mu\text{g}/\text{m}^3$ ); y CDM (1.8, 5.9 y 7.7  $\mu\text{g}/\text{m}^3$ ) tal como se observa en la **Figura 8**.



**Figura 8: Diagrama de cajas y bigotes para las concentraciones de  $\text{SO}_2$**

#### 4.1.1 Patrones espaciales de los contaminantes atmosféricos identificados a través del análisis de componentes principales (ACP)

En la **Tabla 15** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de PM<sub>10</sub>. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, SJL - HCH y STA - ATE, ambos con un valor de 0.71. Las dos correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este del AMLC por lo que comparten características similares de contaminación respecto al PM<sub>10</sub> (Pacsi 2016, Espinoza 2018).

**Tabla 15: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de PM<sub>10</sub>**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.44</b>	1								
CDM	<b>0.40</b>	<b>0.68</b>	1							
STA	<b>0.71</b>	<b>0.61</b>	<b>0.58</b>	1						
VMT	<b>0.33</b>	<b>0.43</b>	<b>0.34</b>	<b>0.46</b>	1					
HCH	<b>0.67</b>	<b>0.45</b>	<b>0.43</b>	<b>0.66</b>	<b>0.39</b>	1				
SJL	<b>0.60</b>	<b>0.46</b>	<b>0.50</b>	<b>0.67</b>	<b>0.39</b>	<b>0.71</b>	1			
SMP	<b>0.22</b>	<b>0.28</b>	<b>0.32</b>	<b>0.25</b>	<b>0.20</b>	<b>0.18</b>	<b>0.23</b>	1		
CRB	<b>0.39</b>	<b>0.42</b>	<b>0.33</b>	<b>0.48</b>	<b>0.47</b>	<b>0.52</b>	<b>0.59</b>	<b>0.16</b>	1	
PPD	<b>0.35</b>	<b>0.40</b>	<b>0.38</b>	<b>0.46</b>	<b>0.48</b>	<b>0.41</b>	<b>0.51</b>	<b>0.18</b>	<b>0.43</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 16** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de PM<sub>2.5</sub>. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, SMP - CDM y HCH - ATE, ambos con un valor de 0.69. Las dos correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona centro y este del AMLC, respectivamente, por lo que presentan comportamientos similares de contaminación respecto al PM<sub>2.5</sub> (Pacsi 2016, Silva *et al.* 2017).

**Tabla 16: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de PM<sub>2,5</sub>**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.35</b>	1								
CDM	<b>0.38</b>	<b>0.55</b>	1							
STA	<b>0.58</b>	<b>0.44</b>	<b>0.43</b>	1						
VMT	<b>0.42</b>	<b>0.38</b>	<b>0.44</b>	<b>0.46</b>	1					
HCH	<b>0.69</b>	<b>0.42</b>	<b>0.47</b>	<b>0.63</b>	<b>0.50</b>	1				
SJL	<b>0.61</b>	<b>0.44</b>	<b>0.49</b>	<b>0.64</b>	<b>0.52</b>	<b>0.66</b>	1			
SMP	<b>0.46</b>	<b>0.54</b>	<b>0.69</b>	<b>0.52</b>	<b>0.49</b>	<b>0.56</b>	<b>0.57</b>	1		
CRB	<b>0.44</b>	<b>0.32</b>	<b>0.43</b>	<b>0.55</b>	<b>0.42</b>	<b>0.53</b>	<b>0.59</b>	<b>0.56</b>	1	
PPD	<b>0.49</b>	<b>0.35</b>	<b>0.44</b>	<b>0.58</b>	<b>0.52</b>	<b>0.56</b>	<b>0.56</b>	<b>0.55</b>	<b>0.59</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 17** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de NO<sub>2</sub>. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, SJL - HCH y SMP - CDM, ambos con valores de 0.83 y 0.68, respectivamente. Estas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Las estaciones antes mencionadas se ubican en la zona este y centro del AMLC, respectivamente, por lo que comparten características similares de contaminación respecto al NO<sub>2</sub> (Pacsi 2016, Sánchez y Ordoñez 2016).

**Tabla 17: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de NO<sub>2</sub>**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.24</b>	1								
CDM	<b>0.13</b>	<b>0.30</b>	1							
STA	<b>-0.39</b>	<b>0.21</b>	<b>0.11</b>	1						
VMT	<b>0.54</b>	<b>-0.05</b>	<b>0.22</b>	<b>-0.58</b>	1					
HCH	<b>0.16</b>	<b>0.45</b>	<b>0.29</b>	<b>0.59</b>	<b>-0.14</b>	1				
SJL	<b>0.27</b>	<b>0.37</b>	<b>0.47</b>	<b>0.36</b>	-0.01	<b>0.83</b>	1			
SMP	<b>0.15</b>	<b>0.39</b>	<b>0.68</b>	<b>0.24</b>	<b>0.09</b>	<b>0.45</b>	<b>0.55</b>	1		
CRB	<b>0.40</b>	<b>0.48</b>	<b>0.35</b>	0.02	<b>0.20</b>	<b>0.49</b>	<b>0.60</b>	<b>0.42</b>	1	
PPD	-0.01	<b>-0.29</b>	<b>0.28</b>	<b>0.04</b>	<b>0.14</b>	<b>0.14</b>	<b>0.28</b>	<b>0.13</b>	<b>0.11</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 18** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de O<sub>3</sub>. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, SJL - STA y STA - ATE, con valores de 0.92 y 0.90, respectivamente. Las dos correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este del AMLC, por lo que presentan comportamientos similares de contaminación respecto al O<sub>3</sub> (Pacsi 2016, Espinoza 2018).

**Tabla 18: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de O<sub>3</sub>**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.60</b>	1								
CDM	<b>0.63</b>	<b>0.78</b>	1							
STA	<b>0.90</b>	<b>0.67</b>	<b>0.70</b>	1						
VMT	<b>0.70</b>	<b>0.68</b>	<b>0.65</b>	<b>0.80</b>	1					
HCH	<b>0.83</b>	<b>0.54</b>	<b>0.58</b>	<b>0.87</b>	<b>0.71</b>	1				
SJL	<b>0.87</b>	<b>0.62</b>	<b>0.65</b>	<b>0.92</b>	<b>0.76</b>	<b>0.84</b>	1			
SMP	<b>0.63</b>	<b>0.67</b>	<b>0.68</b>	<b>0.67</b>	<b>0.67</b>	<b>0.58</b>	<b>0.66</b>	1		
CRB	<b>0.70</b>	<b>0.44</b>	<b>0.51</b>	<b>0.76</b>	<b>0.61</b>	<b>0.79</b>	<b>0.77</b>	<b>0.62</b>	1	
PPD	<b>0.80</b>	<b>0.49</b>	<b>0.54</b>	<b>0.78</b>	<b>0.62</b>	<b>0.75</b>	<b>0.76</b>	<b>0.68</b>	<b>0.80</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 19** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de CO. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, SJL - HCH y CDM - SBJ, con valores de 0.81 y 0.71, respectivamente. Estas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Las estaciones antes mencionadas se ubican en la zona este y centro del AMLC, por lo que comparten características similares de contaminación respecto al CO (Pacsi 2016, Espinoza y Alderete 2021).

En la **Tabla 20** se presenta la matriz de correlación de Pearson entre las estaciones de monitoreo sobre los datos de las concentraciones de SO<sub>2</sub>. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de, STA - ATE y SJL - STA, con valores de 0.55 y 0.54, respectivamente. Las dos correlaciones son estadísticamente

significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este del AMLC, por lo que presentan comportamientos similares de contaminación respecto al SO<sub>2</sub> (Pacsi 2016, Sánchez y Ordoñez 2016).

**Tabla 19: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de CO**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.15</b>	1								
CDM	<b>0.10</b>	<b>0.71</b>	1							
STA	<b>0.53</b>	<b>0.55</b>	<b>0.46</b>	1						
VMT	<b>0.55</b>	<b>0.31</b>	<b>0.51</b>	<b>0.51</b>	1					
HCH	<b>0.54</b>	<b>0.30</b>	<b>0.43</b>	<b>0.55</b>	<b>0.69</b>	1				
SJL	<b>0.38</b>	<b>0.40</b>	<b>0.58</b>	<b>0.62</b>	<b>0.63</b>	<b>0.81</b>	1			
SMP	<b>-0.11</b>	<b>0.57</b>	<b>0.35</b>	<b>0.47</b>	<b>-0.11</b>	0.01	<b>0.23</b>	1		
CRB	<b>0.36</b>	<b>0.14</b>	<b>0.23</b>	<b>0.36</b>	<b>0.44</b>	<b>0.59</b>	<b>0.58</b>	<b>-0.04</b>	1	
PPD	<b>0.51</b>	<b>-0.03</b>	0.02	<b>0.19</b>	<b>0.39</b>	<b>0.29</b>	<b>0.23</b>	<b>-0.27</b>	<b>0.53</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

**Tabla 20: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de concentraciones de SO<sub>2</sub>**

Estaciones	ATE	SBJ	CDM	STA	VMT	HCH	SJL	SMP	CRB	PPD
ATE	1									
SBJ	<b>0.31</b>	1								
CDM	<b>0.16</b>	<b>0.44</b>	1							
STA	<b>0.55</b>	<b>0.27</b>	<b>0.29</b>	1						
VMT	<b>0.23</b>	<b>0.23</b>	<b>0.08</b>	<b>0.17</b>	1					
HCH	<b>0.44</b>	<b>0.24</b>	<b>0.09</b>	<b>0.44</b>	<b>0.22</b>	1				
SJL	<b>0.54</b>	<b>0.37</b>	<b>0.34</b>	<b>0.54</b>	<b>0.32</b>	<b>0.29</b>	1			
SMP	<b>0.31</b>	<b>0.30</b>	<b>0.40</b>	<b>0.37</b>	0.06	<b>0.21</b>	<b>0.41</b>	1		
CRB	0.06	<b>0.11</b>	<b>0.16</b>	<b>0.12</b>	-0.06	0.02	<b>0.37</b>	<b>0.28</b>	1	
PPD	<b>0.22</b>	<b>0.27</b>	<b>0.10</b>	<b>0.13</b>	<b>0.35</b>	<b>0.10</b>	<b>0.34</b>	<b>0.19</b>	0.05	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 21** se presentan los resultados de las pruebas de Kaiser-Meyer-Olkin (KMO) y esfericidad de Bartlett de las concentraciones de contaminantes para las diez estaciones de

monitoreo de calidad de aire. Se observa que, los valores de medida de KMO son mayores a 0.5, lo que indica que los registros de los contaminantes atmosféricos son adecuados para el análisis estadístico multivariado (Abdullah *et al.* 2018). Así también, la significancia de las pruebas de esfericidad de Bartlett son inferiores al 5%, lo que refuerza que, las concentraciones de los contaminantes están correlacionadas entre sí, por lo tanto, es factible realizar el análisis de componentes principales (ACP) (Ul-Saufie *et al.* 2013, Nazmi *et al.* 2020).

**Tabla 21: Pruebas de KMO y esfericidad de Bartlett de los contaminantes atmosféricos**

Prueba de KMO y Bartlett	PM <sub>10</sub>	PM <sub>2,5</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	SO <sub>2</sub>
Medida Kaiser-Meyer-Olkin	0.887	0.926	0.690	0.925	0.766	0.773
Prueba de Aprox. Chi-cuadrado	19232.0	6938.4	12650.1	17463.8	24676.4	1661.2
esfericidad	gl	45	45	45	45	45
de Bartlett	Sig.	0.000	0.000	0.000	0.000	0.000

Se realizó el ACP utilizando los datos de concentraciones de PM<sub>10</sub>, PM<sub>2,5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO; y SO<sub>2</sub>, medidos en las diez estaciones que conforman la Red de Monitoreo Automático de la Calidad del Aire (REMCA) del AMLC. La **Tabla 22** resume los resultados principales de la aplicación del ACP en todas las estaciones para el contaminante PM<sub>10</sub>. El análisis extrajo un total de seis CP para este contaminante. Basándose únicamente en las CP que cumplían con el criterio de Kaiser, las primeras CP del PM<sub>10</sub> explicaron el 60.8% de la varianza total; por lo que, fue necesario extraer cuatro CP adicionales con la finalidad de cumplir con el segundo criterio, es decir, explicar al menos el 85% de la varianza total para el PM<sub>10</sub> (Zhao *et al.* 2015, Wang *et al.* 2018).

Los resultados del ACP para el contaminante PM<sub>2,5</sub> se presentan en la **Tabla 23**, en esta se puede ver que, la primera CP explica, el 55.9% de la varianza de los datos originales de las concentraciones de PM<sub>2,5</sub>, seguidas de las componentes CP2, CP3, CP4, CP5 y CP6 con varianzas acumuladas de 65.5%, 72.4%, 78.3%, 83.2% y 87.3%, respectivamente. Los factores de carga rotados demarcados en negrita indican las variables que más influyen en la CP correspondiente. En la primera CP (CP1), las estaciones de ATE, STA, HCH y SJL parecen presentar características similares y se combinan para asumir mucha carga al describir el nivel de contaminación de PM<sub>2,5</sub>. La CP2 está fuertemente cargada por la contribución de las estaciones CDM y SBJ; mientras que, la CP3 se llena principalmente por la contribución de las estaciones CRB y SJL. Las tres últimas CP se ven influenciadas por



una sola estación individualmente, es decir las estaciones SBJ, VMT y PPD contribuyen a las CP4, CP5 y CP6, respectivamente. Dado que estos valores de carga correspondientes son todos superiores a 0.5, se puede juzgar que estas estaciones podrían agruparse claramente en diferentes categorías en las que la variación del PM<sub>2,5</sub> en estas estaciones muestra un comportamiento similar (Pires *et al.* 2008, Lu *et al.* 2011).

**Tabla 22: Principales resultados de la aplicación del ACP a los datos de concentraciones de PM<sub>10</sub>**

	CP1	CP2	CP3	CP4	CP5	CP6
ATE	<b>0.89</b>	0.17	0.03	0.06	0.14	0.10
SBJ	0.25	<b>0.82</b>	0.17	0.08	0.24	0.08
CDM	0.25	<b>0.85</b>	0.10	0.18	0.02	0.16
STA	<b>0.71</b>	0.45	0.14	0.16	0.22	0.07
VMT	0.20	0.18	0.21	0.21	<b>0.89</b>	0.08
HCH	<b>0.79</b>	0.18	0.34	0.15	0.09	0.03
SJL	<b>0.64</b>	0.25	0.47	0.35	-0.02	0.09
SMP	0.10	0.16	0.04	0.05	0.07	<b>0.98</b>
CRB	0.25	0.18	<b>0.88</b>	0.13	0.24	0.04
PPD	0.21	0.19	0.15	<b>0.90</b>	0.22	0.06
Autovalores	5.05	1.03	0.90	0.75	0.57	0.55
% de varianza	50.5	10.3	9.0	7.5	5.7	5.5
% acumulado	50.5	60.8	69.8	77.4	83.1	88.6

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

**Tabla 23: Principales resultados de la aplicación del ACP a los datos de concentraciones de PM<sub>2,5</sub>**

	CP1	CP2	CP3	CP4	CP5	CP6
ATE	<b>0.88</b>	0.18	0.11	0.09	0.11	0.16
SBJ	0.16	0.36	0.08	<b>0.87</b>	0.13	0.07
CDM	0.20	<b>0.85</b>	0.13	0.24	0.15	0.10
STA	<b>0.53</b>	0.05	0.45	0.42	0.13	0.32
VMT	0.23	0.22	0.16	0.14	<b>0.90</b>	0.20
HCH	<b>0.75</b>	0.27	0.24	0.13	0.20	0.21
SJL	<b>0.56</b>	0.23	<b>0.56</b>	0.22	0.29	0.01
SMP	0.26	<b>0.73</b>	0.32	0.22	0.17	0.21
CRB	0.20	0.27	<b>0.83</b>	0.04	0.11	0.28
PPD	0.29	0.23	0.28	0.10	0.23	<b>0.82</b>
Autovalores	5.59	0.96	0.68	0.60	0.49	0.41
% de varianza	55.9	9.6	6.8	6.0	4.9	4.1
% acumulado	55.9	65.5	72.4	78.3	83.2	87.3

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Para el contaminante NO<sub>2</sub>, se puede observar en la **Tabla 24** que, las tres primeras CP explican el 72.3% de la varianza de los datos originales de las concentraciones de NO<sub>2</sub>, seguidas de las componentes CP4 y CP5 con varianzas acumuladas de 81.7% y 86.8%, respectivamente. En la primera CP (CP1), las estaciones HCH, SJL y STA parecen presentar comportamientos similares y se combinan para asumir mucha carga al describir el nivel de contaminación de NO<sub>2</sub>. La CP2 está fuertemente cargada por la contribución de las estaciones ATE y VMT; mientras que, la CP3 se llena principalmente por la contribución de las estaciones CDM y SMP. Los componentes restantes se cargan por la contribución de las estaciones CRB y SBJ (CP4); y PPD (CP5). Dado que estos valores de carga correspondientes son todos superiores a 0.5, se puede juzgar que, estas estaciones podrían agruparse claramente en diferentes categorías, en las que la variación de NO<sub>2</sub> en estos sitios de monitoreo muestren un comportamiento similar (Pires *et al.* 2008, Wang *et al.* 2018).

**Tabla 24: Principales resultados de la aplicación del ACP a los datos de concentraciones de NO<sub>2</sub>**

	CP1	CP2	CP3	CP4	CP5
ATE	0.22	<b>0.86</b>	-0.01	0.21	-0.13
SBJ	0.24	-0.02	0.31	<b>0.61</b>	-0.55
CDM	0.09	0.08	<b>0.90</b>	0.16	0.16
STA	<b>0.66</b>	-0.64	0.13	-0.10	-0.07
VMT	-0.19	<b>0.85</b>	0.19	-0.01	0.14
HCH	<b>0.91</b>	-0.05	0.18	0.27	-0.01
SJL	<b>0.77</b>	0.10	0.32	0.37	0.19
SMP	0.32	0.05	<b>0.84</b>	0.15	-0.05
CRB	0.28	0.22	0.19	<b>0.86</b>	0.08
PPD	0.12	0.02	0.15	0.04	<b>0.93</b>
Autovalores	3.68	2.18	1.37	0.94	0.51
% de varianza	36.8	21.8	13.7	9.4	5.1
% acumulado	36.8	58.6	72.3	81.7	86.8

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Para el contaminante O<sub>3</sub>, se puede ver en la **Tabla 25** que, la primera CP explica el 72.9% de la varianza de los datos originales de las concentraciones de O<sub>3</sub>, seguidas de las componentes CP2 y CP3 con varianzas acumuladas de 82.3% y 87.1%, respectivamente. En

la primera CP (CP1), las estaciones HCH, STA, SJL y ATE, parecen presentar características similares y se combinan para asumir mucha carga al describir el nivel de contaminación de O<sub>3</sub>. La CP2 está fuertemente cargada por la contribución de las estaciones SBJ, CDM, SMP y VMT; mientras que, la CP3 se ve influenciada por las estaciones PPD y CRB (Wang *et al.* 2018).

**Tabla 25: Principales resultados de la aplicación del ACP a los datos de concentraciones de O<sub>3</sub>**

	CP1	CP2	CP3
ATE	<b>0.80</b>	0.37	0.29
SBJ	0.29	<b>0.88</b>	0.15
CDM	0.32	<b>0.82</b>	0.24
STA	<b>0.81</b>	0.46	0.27
VMT	0.61	<b>0.61</b>	0.17
HCH	<b>0.85</b>	0.27	0.30
SJL	<b>0.80</b>	0.40	0.30
SMP	0.23	<b>0.62</b>	0.68
CRB	0.65	0.15	<b>0.64</b>
PPD	0.62	0.22	<b>0.67</b>
Autovalores	7.29	0.94	0.48
% de varianza	72.9	9.4	4.8
% acumulado	72.9	82.3	87.1

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Para el contaminante CO, de la **Tabla 26** se observa que, las dos primeras CP explican el 66.8% de la varianza de los datos originales de las concentraciones de CO; seguidas de las componentes CP3, CP4 y CP5 con varianzas acumuladas de 75.3%, 82.7% y 89.4%, respectivamente. Los factores de carga rotados en negrita indican las variables que más influyen en la CP correspondiente. En la primera CP (CP1), las estaciones SJL, HCH, y VMT parecen presentar comportamientos similares y se combinan para asumir mucha carga al describir el nivel de contaminación de CO. La CP2 está fuertemente cargada por la contribución de las estaciones CDM y SBJ; mientras que, la CP3 se llena principalmente por la contribución de las estaciones ATE y STA. Los componentes restantes se cargan por la

contribución de sus propias estaciones, respectivamente, es decir, CP4 (SMP) y CP5 (PPD y CRB). Dado que, estos valores de carga correspondientes son todos superiores a 0.5, se puede juzgar que, estas estaciones podrían agruparse claramente en diferentes categorías en las que la variación de CO en estas estaciones muestra un comportamiento similar (Lu *et al.* 2011).

**Tabla 26: Principales resultados de la aplicación del ACP a los datos de concentraciones de CO**

	CP1	CP2	CP3	CP4	CP5
ATE	0.23	-0.01	<b>0.88</b>	-0.01	0.27
SBJ	0.08	<b>0.76</b>	0.13	0.51	0.02
CDM	0.33	<b>0.88</b>	-0.01	0.16	0.00
STA	0.42	0.25	<b>0.56</b>	0.56	0.07
VMT	<b>0.54</b>	0.48	0.50	-0.26	0.14
HCH	<b>0.85</b>	0.19	0.36	-0.02	0.09
SJL	<b>0.85</b>	0.31	0.18	0.18	0.09
SMP	0.02	0.23	-0.08	<b>0.91</b>	-0.14
CRB	0.67	0.00	-0.02	0.08	<b>0.67</b>
PPD	0.06	0.04	0.33	-0.17	<b>0.88</b>
Autovalores	4.54	2.14	0.85	0.74	0.67
% de varianza	45.4	21.4	8.5	7.4	6.7
% acumulado	45.4	66.8	75.3	82.7	89.4

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Finalmente, de los resultados del ACP para el SO<sub>2</sub> presentados en la **Tabla 27**, se puede observar que, las tres primeras CP explican el 59.5% de la varianza de los datos originales de las concentraciones de SO<sub>2</sub>, seguidas de las componentes CP4, CP5, CP6 y CP7 con varianzas acumuladas de 69.1%, 75.9%, 82.2% y 88.2%, respectivamente. En la primera CP (CP1), las estaciones ATE, STA y SJL parecen presentar características similares y se combinan para asumir mucha carga al describir el nivel de contaminación de SO<sub>2</sub>. La CP2 está fuertemente cargada por la contribución de las estaciones SBJ y CDM; mientras que, la CP3 se llena principalmente por la contribución de la estación CRB. Las componentes restantes se cargan por la contribución de sus propias estaciones, respectivamente, es decir, CP4 (SMP), CP5 (PPD), CP6 (VMT) y CP7 (HCH) (Zhao *et al.* 2015).

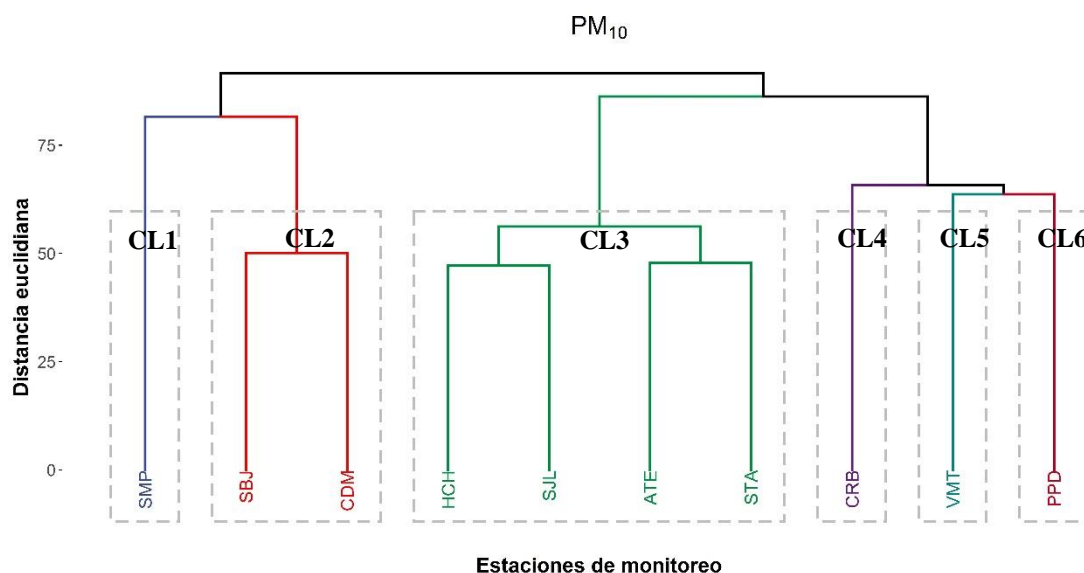
**Tabla 27: Principales resultados de la aplicación del ACP a los datos de concentraciones de SO<sub>2</sub>**

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
ATE	<b>0.83</b>	0.10	-0.03	0.02	0.18	0.00	0.24
SBJ	0.16	<b>0.87</b>	0.07	-0.01	0.24	0.04	0.21
CDM	0.16	<b>0.71</b>	0.04	0.48	-0.17	0.12	-0.19
STA	<b>0.80</b>	0.11	0.01	0.24	-0.10	0.08	0.18
VMT	0.13	0.08	-0.04	-0.02	0.19	<b>0.95</b>	0.11
HCH	0.31	0.08	0.00	0.10	-0.02	0.12	<b>0.90</b>
SJL	<b>0.69</b>	0.21	0.40	0.16	0.21	0.25	-0.03
SMP	0.22	0.13	0.16	<b>0.88</b>	0.15	-0.04	0.13
CRB	0.06	0.05	<b>0.97</b>	0.13	0.00	-0.05	0.00
PPD	0.12	0.10	0.02	0.11	<b>0.92</b>	0.20	-0.01
Autovalores	3.46	1.34	1.16	0.96	0.68	0.63	0.60
% de varianza	34.6	13.4	11.6	9.6	6.8	6.3	6.0
% acumulado	34.6	47.9	59.5	69.1	75.9	82.2	88.2

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

#### 4.1.2 Patrones espaciales de los contaminantes atmosféricos identificados a través del análisis de clúster jerárquico (ACJ)

El análisis de clúster jerárquico (ACJ) se usó para agrupar estaciones de monitoreo basados en la similitud de los datos de concentraciones máxicas estandarizadas de PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO y SO<sub>2</sub>. La **Figura 9** presenta los resultados del método de ACJ sobre el contaminante PM<sub>10</sub>. De acuerdo con el valor de la distancia euclidiana entre las estaciones de monitoreo del PM<sub>10</sub> calculado en el software R, las diez estaciones de la REMCA pueden ser agrupados en seis grupos: clúster I (CL1)-estación SMP; clúster II (CL2)-estaciones SBJ y CDM; clúster III (CL3)-estaciones HCH, SJL, ATE y STA; clúster IV (CL4)-estación CRB; clúster V (CL5)-estación VMT y clúster VI (CL6)-estación PPD. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 22**), donde los componentes principales CP1, CP2, CP3, CP4, CP5 y CP6 corresponden a los grupos CL3, CL2, CL4, CL6, CL5 y CL1 del ACJ, respectivamente.

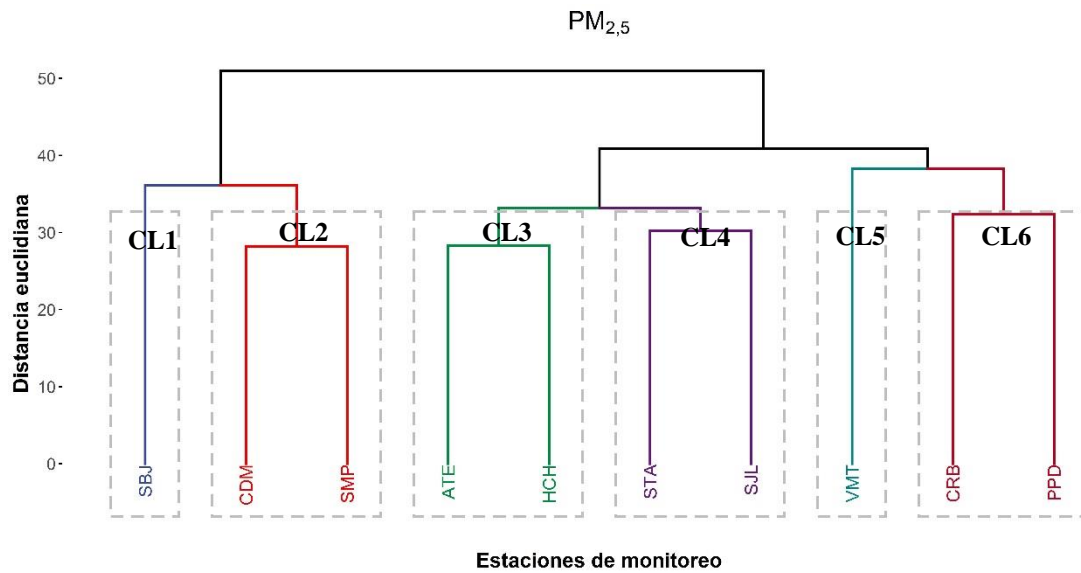


**Figura 9: Dendrograma resultante de la aplicación del ACJ a las concentraciones de  $PM_{10}$**

Se puede encontrar que las estaciones agrupadas dependen evidentemente de la geografía, y las fuentes de emisión particulares de las zonas. Las estaciones en el grupo CL3 se encuentran en la zona este de Lima Metropolitana, las cuales están influenciadas principalmente por el desarrollo industrial, alto tráfico vehicular, mala pavimentación de sus avenidas y ausencia de áreas verdes; mientras que, las estaciones en el grupo CL2 se encuentran en la zona centro de Lima Metropolitana, que presentan características urbanísticas y meteorológicas propias de la zona, como las vías en buen estado y mayores áreas verdes (Pacsi, 2016; MINAM, 2016). Tanto la geografía y las fuentes de emisión locales son factores que juegan un papel importante en la prevención de la dispersión de contaminantes y dan como resultado esta clasificación.

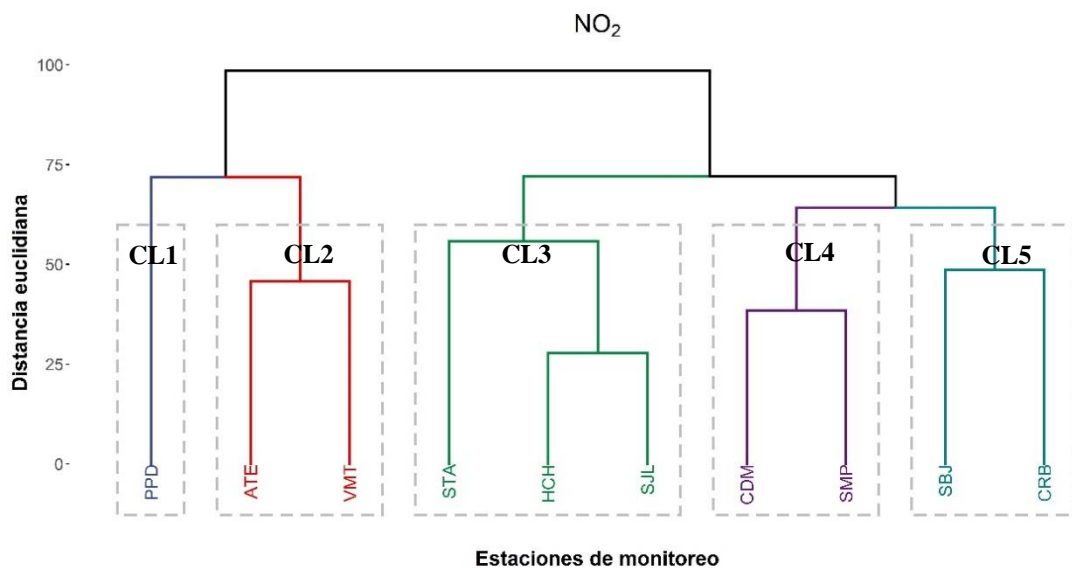
De manera similar, el dendrograma resultante de la aplicación del ACJ en las concentraciones máxicas de  $PM_{2,5}$  se muestra en la **Figura 10**. Se puede encontrar que, las estaciones de monitoreo podrían agruparse en diferentes categorías en términos de la distancia euclidiana. Para las concentraciones máxicas de  $PM_{2,5}$ , las diez estaciones de la REMCA pueden ser agrupados en seis grupos: clúster I (CL1)-estación SBJ, clúster II (CL2)-estaciones CDM y SMP, clúster III (CL3)-estaciones ATE y HCH; clúster IV (CL4)-estaciones STA, y SJL; clúster V (CL5)-estación VMT y clúster VI (CL6)-estaciones CRB y PPD. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 23**), donde los CP1, CP2, CP3, CP4, CP5 y el CP6, correspondieron a los grupos CL3 y CL4, CL2, CL6, CL1, CL5 y CL6 del ACJ, respectivamente. También se puede concluir que, los

grupos de estas estaciones dependen en gran medida de la geografía y las fuentes de emisión del  $PM_{2,5}$ , tanto del parque automotor e industrial (Silva *et al.* 2017).



**Figura 10: Dendrograma resultante de la aplicación del ACJ a las concentraciones de  $PM_{2,5}$**

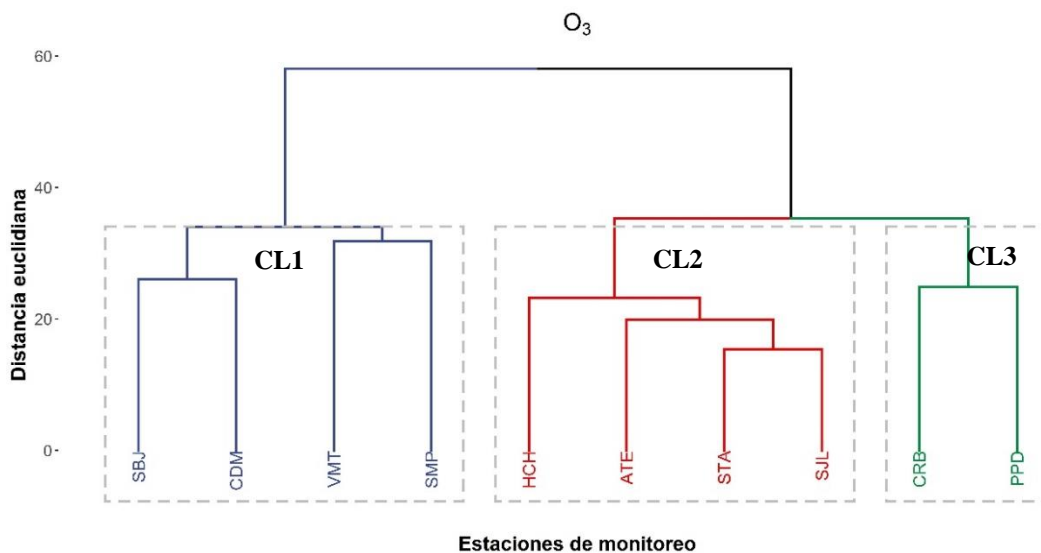
La **Figura 11** muestra los resultados del método del ACJ sobre el contaminante  $NO_2$ . De acuerdo con el valor de la distancia euclidiana entre las estaciones de monitoreo del  $NO_2$  calculado en el software R, las diez estaciones de la REMCA pueden ser agrupados en cinco grupos: clúster I (CL1)-estación PPD; clúster II (CL2)-estaciones ATE y VMT; clúster III (CL3)-estaciones STA, HCH y SJL; clúster IV (CL4)-estaciones CDM y SMP; y clúster V (CL5)-estaciones SBJ y CRB. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 24**), donde los componentes principales CP1, CP2, CP3, CP4 y CP5 corresponden a los grupos CL3, CL2, CL4, CL5 y CL1 del ACJ, respectivamente.



**Figura 11: Dendrograma resultante de la aplicación del ACJ a las concentraciones de  $NO_2$**

Se puede encontrar que, estos grupos dependen de su ubicación geográfica, las condiciones meteorológicas y las fuentes de emisión propias de las zonas de estudio (Gramsch *et al.* 2006). Las estaciones del grupo CL3 se encuentran en la zona este de Lima Metropolitana y presentan los mayores niveles de NO<sub>2</sub> en promedio, siendo la estación HCH, la de mayor concentración promedio dentro del grupo; así también, las estaciones del grupo CL2 se encuentran en la zona este y sur de Lima Metropolitana, y se caracterizan por presentar los mayores y menores niveles promedio de NO<sub>2</sub>, siendo la estación ATE la de mayores concentraciones del contaminante entre todas las estaciones de la REMCA, esto se debe a su ubicación geográfica a favor del viento, es decir, se localiza en una zona angosta de la microcuenca del río Rímac, donde los cerros obstaculizan la dispersión de los contaminantes atmosféricos (Pacsi 2016, MINAM 2019). Sumado al transporte del NO<sub>2</sub> desde la zona sur hacia la zona este, debido a la dirección predominante de los vientos en el AMLC (Silva *et al.* 2017).

De manera similar, el dendrograma resultante de la aplicación del ACJ en las concentraciones máxicas de O<sub>3</sub> se muestra en la **Figura 12**. Se puede encontrar que, las estaciones de monitoreo podrían agruparse en diferentes categorías en términos de la distancia euclidiana. Para las concentraciones máxicas de O<sub>3</sub>, las diez estaciones de la REMCA pueden ser agrupados en tres grupos: clúster I (CL1)-estaciones SBJ, CDM, VMT y SMP, clúster II (CL2)-estaciones HCH, ATE, STA y SJL; y clúster III (CL3)-estaciones CRB y PPD. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 25**), donde los CP1, CP2 y CP3, corresponden a los grupos CL2, CL1 y CL3 del ACJ, respectivamente.

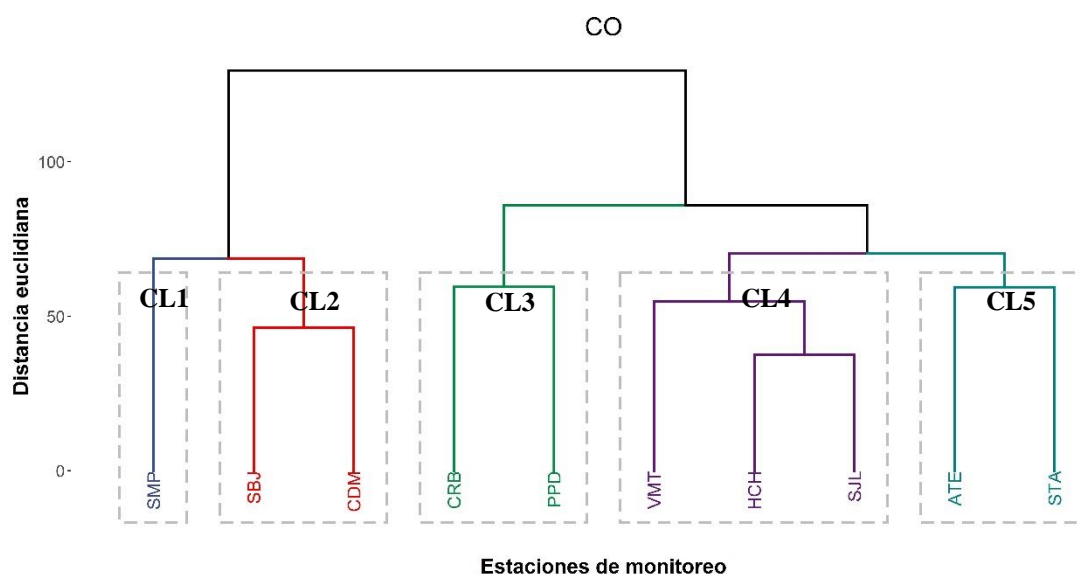


**Figura 12: Dendrograma resultante de la aplicación del ACJ a las concentraciones de O<sub>3</sub>**



Las estaciones en el grupo CL1 se encuentran en la zona centro y sur de Lima Metropolitana; mientras que, las estaciones en el grupo CL2 y CL3 se encuentran en la zona este y norte del área de estudio, respectivamente, las cuales están influenciadas principalmente por el desarrollo industrial, alto tráfico vehicular y ausencia de áreas verdes (Pacsi 2016, MINAM 2016). El CL1 presenta las mayores concentraciones de O<sub>3</sub>, el CL3, intermedias y el CL2, menores concentraciones. Se observa una tendencia contraria frente a la distribución del NO<sub>2</sub>, es decir las estaciones con mayores niveles de NO<sub>2</sub> presentan los menores valores de O<sub>3</sub>, lo cual puede deberse a un proceso de trituración (Silva *et al.* 2018). Se puede encontrar que, tanto la geografía, la meteorología y las fuentes de emisión locales son las que juegan un papel importante en la dispersión del O<sub>3</sub> y dan como resultado esta clasificación (Pires *et al.* 2008).

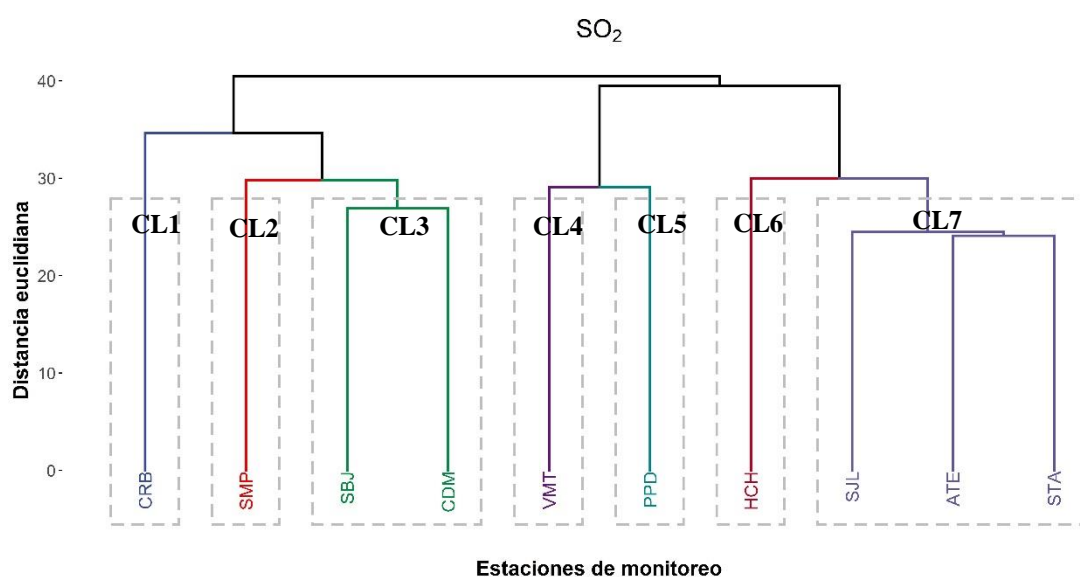
La **Figura 13** muestra los resultados del método del ACJ sobre el contaminante CO. De acuerdo con el valor de la distancia euclidiana del CO calculado en el software R, las diez estaciones de la REMCA pueden ser agrupados en cinco grupos: clúster I (CL1)-estación SMP; clúster II (CL2)-estaciones SBJ y CDM; clúster III (CL3)-estaciones CRB y PPD; clúster IV (CL4)-estaciones VMT, HCH y SJL; y clúster V (CL5)-estaciones ATE y STA. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 26**), donde los componentes principales CP1, CP2, CP3, CP4 y CP5 corresponden a los grupos CL4, CL2, CL5, CL1 y CL3 del ACJ, respectivamente.



**Figura 13: Dendrograma resultante de la aplicación del ACJ a las concentraciones de CO**

Se puede observar que, estos grupos dependen de la geografía, las condiciones meteorológicas y las fuentes de emisión particulares de las zonas de estudio (Espinoza y Alderete 2021). Las estaciones en el grupo CL2 se encuentran en la zona centro de Lima Metropolitana, esta región geográfica está rodeada por áreas verdes y presenta características urbanísticas y meteorológicas particulares, como las vías en buen estado y mayores áreas verdes; mientras que, las estaciones en los grupos CL4 y CL5 se encuentran en la zona este y sur de Lima Metropolitana, las cuales están influenciadas principalmente por el desarrollo industrial, alto tráfico vehicular, mala pavimentación de sus avenidas y ausencia de áreas verdes (Pacsi 2016, MINAM 2019). Tanto la geografía y las fuentes de emisión locales son factores que juegan un papel importante en la prevención de la dispersión de contaminantes y dan como resultado esta clasificación para el contaminante CO.

La **Figura 14** muestra los resultados del método del ACJ sobre el contaminante SO<sub>2</sub>. De acuerdo con el valor de la distancia euclidiana del SO<sub>2</sub> calculado en el software R, las diez estaciones de la REMCA pueden ser agrupados en siete grupos: clúster I (CL1)-estación CRB; clúster II (CL2)-estación SMP; clúster III (CL3)-estaciones SBJ y CDM; clúster IV (CL4)-estación VMT; clúster V (CL5)-estación PPD; clúster VI (CL6)-estación HCH y clúster VII (CL7)-estaciones SJL, ATE y STA. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 27**), donde los componentes principales CP1, CP2, CP3, CP4, CP5, CP6 y CP7 correspondían a los grupos CL7, CL3, CL1, CL2, CL5, CL4 y CL6 del ACJ, respectivamente.



**Figura 14: Dendrograma resultante de la aplicación del ACJ a las concentraciones de SO<sub>2</sub>**

Se puede encontrar que, estos grupos dependen de la geografía, las características meteorológicas y las fuentes de emisión de las zonas de estudio. Las estaciones en el grupo CL3 se encuentran en la zona centro de Lima Metropolitana, la cual está rodeada por áreas verdes y presenta vías en buen estado, por ende evidencia las menores concentraciones de SO<sub>2</sub>; mientras que, las estaciones en los grupos CL6 y CL7 se encuentran en la zona este de Lima Metropolitana, las cuales están influenciadas principalmente por el desarrollo industrial, alto tráfico vehicular y ausencia de áreas verdes, presentando de esta manera los mayores niveles de contaminación por SO<sub>2</sub> (**Tabla 14**) (Espinoza 2018, MINAM 2019). Tanto la geografía y las fuentes de emisión locales son factores que influyen en la prevención de la dispersión del SO<sub>2</sub> y dan como resultado esta categorización.

#### **4.2 DETERMINACIÓN DE LOS PATRONES ESPACIALES DE LAS VARIABLES METEOROLÓGICAS MEDIDOS EN LA RED DE ESTACIONES METEOROLÓGICAS AUTOMÁTICAS DEL ÁREA METROPOLITANA DE LIMA MEDIANTE EL ACP Y EL ACJ**

Se aplicaron estadísticas descriptivas sobre los datos de las variables meteorológicas (temperatura del aire, humedad relativa, dirección, velocidad del viento y radiación solar) a nivel horario para comprender la variabilidad de los parámetros meteorológicos en el área Metropolitana de Lima. Los resultados de la estadística descriptiva se presentan en la **Tabla 28**, éstos incluyen el máximo, el mínimo y la media de las variables meteorológicas por cada estación de monitoreo.

En promedio, la temperatura del aire varía entre 18.3 y 20.2 °C; la humedad relativa, 75.2 y 87.2 %; la dirección del viento, 191.4 y 248.4 °; la velocidad del viento, 1.1 y 2.2 m/s; y la radiación solar, 136.9 y 170.1 Wh/m<sup>2</sup>. Los valores máximos de temperatura del aire se alcanzan en la estación PPD (33.1°C); y los valores mínimos, en la estación SMP (10.6°C). Para la humedad relativa el valor máximo se presenta en la estación STA y SBJ (100%) y el mínimo se presenta en CRB (31.5%). De la misma manera, la radiación solar presenta sus valores máximos en la estación HCH (1257.0 Wh/m<sup>2</sup>). En el caso de la velocidad del viento los valores máximos se presentan en la estación CDM (8.5 m/s), mientras los mínimos se alcanzan en hasta 5 estaciones.

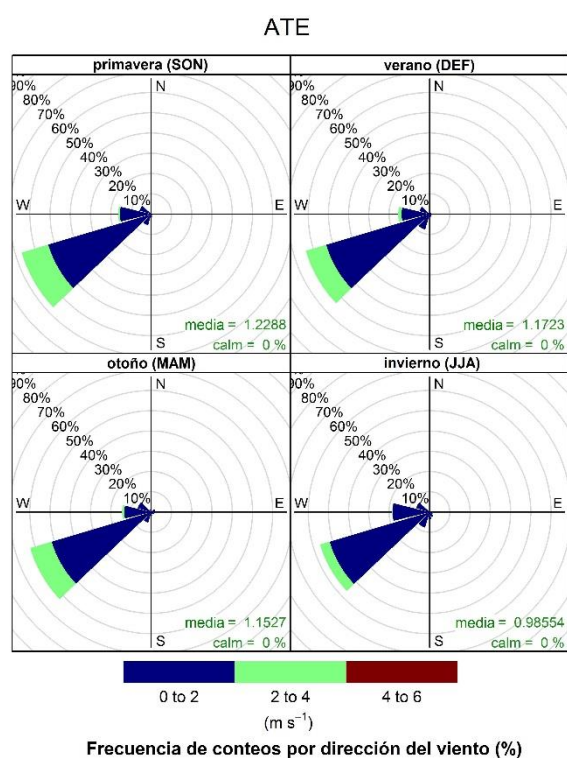
**Tabla 28: Estadísticas descriptivas de las variables meteorológicas**

Estación	ATE	STA	SMP	SJL	SBJ	PPD	HCH	CRB	CDM	VMT	
T	Min	12.3	12.5	10.6	12.6	13.0	12.9	12.9	13.4	13.2	12.3
	Media	19.3±0.0	18.3±0.0	18.6±0.0	19.8±0.0	19.0±0.0	19.6±0.0	19.3±0.0	20.2±0.0	19.5±0.0	19.3±0.0
	Max	32.0	31.8	31.7	32.1	32.2	33.1	31.4	32.9	31.0	31.0
	DE	3.9	3.6	3.6	3.9	3.5	3.9	4.0	3.9	3.3	3.7
HR	Min	35.3	37.0	41.4	35.2	45.0	35.1	37.2	31.5	42.0	38.1
	Media	80.9±0.1	84.2±0.1	79.9±0.1	75.2±0.1	87.2±0.1	78.5±0.1	80.7±0.1	78.2±0.1	83.0±0.0	85.6±0.1
	Max	99.8	100.0	99.8	98.9	100.0	99.6	99.4	99.7	98.0	99.8
	DE	12.5	10.8	8.7	11.7	10.0	11.8	12.7	11.8	8.1	11.9
DV	Min	21.3	1.0	7.1	16.6	1.0	4.6	9.1	18.3	1.0	1.1
	Media	239.6±0.3	205.9±0.3	191.4±0.3	202.9±0.2	202.9±0.3	219.2±0.5	248.4±0.4	199.4±0.2	217.8±0.2	197.9±0.6
	Max	347.2	360.0	326.3	351.4	360.0	350.6	349.6	355.6	360	342.7
	DE	39.3	37.1	39.5	27.5	41.8	67.0	42.3	29.6	35.2	63.2
VV	Min	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0
	Media	1.1±0.0	1.7±0.0	2.0±0.0	1.8±0.0	1.4±0.0	1.4±0.0	1.9±0.0	1.5±0.0	2.2±0.0	1.2±0.0
	Max	3.7	5.7	5.4	5.1	4.8	5.3	5.9	5.2	8.5	5.3
	DE	0.6	0.9	0.8	1.0	0.7	0.8	1.1	0.8	1.2	0.7
RS	Min	0.0	N.R.	0.0	0.0	N.R.	0.0	0.0	0.0	N.R.	0.0
	Media	164.5±1.7	N.R.	159.9±1.6	157.1±1.7	N.R.	136.9±1.5	156.2±2.1	138.5±1.6	N.R.	170.1±2.4
	Max	1189.0	N.R.	1250.0	1121.0	N.R.	1184.0	1257.0	1021.0	N.R.	1246.0
	DE	260.8	N.R.	257.0	251.8	N.R.	224.4	250.3	223.3	N.R.	285.7

DE: desviación estándar; N.R.: no registra

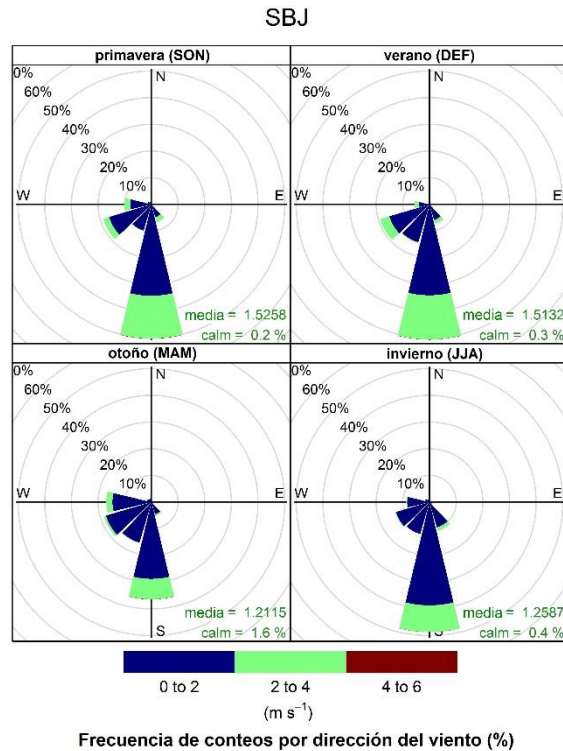
En las **Figuras 15** al **24** se presentan las rosas de vientos para las diez estaciones de monitoreo a nivel estacional para el periodo 2015-2019. Las estaciones comprenden primavera (setiembre, octubre y noviembre), verano (diciembre, enero y febrero), otoño (marzo, abril y mayo) e invierno (junio, julio y agosto).

Las rosas de vientos para la estación ATE se presentan en la **Figura 15**. Se evidencia que, la dirección predominante en las cuatro estaciones del año es la del suroeste (SO), variando su velocidad en promedio de 1.22, 1.17, 1.15 y 0.98 m/s en las estaciones de primavera, verano, otoño e invierno, respectivamente. Estas características influyen en la dispersión de los contaminantes atmosféricos y por ende en la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).



**Figura 15: Rosa de vientos para la estación ATE, periodo 2015-2019**

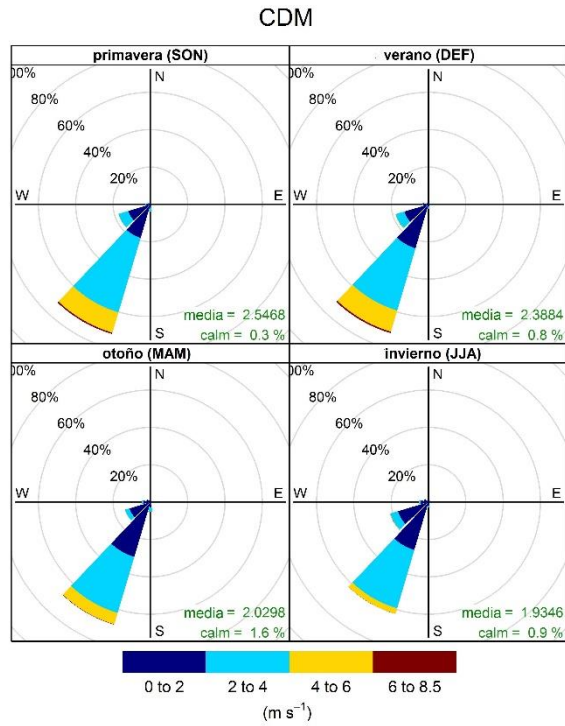
Las rosas de vientos para la estación SBJ se presentan en la **Figura 16**. La dirección predominante del viento en las cuatro estaciones del año es la del sur (S), variando su velocidad en promedio de 1.52, 1.51, 1.21 y 1.26 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Estas variaciones influyen en el transporte de los contaminantes atmosféricos y, por ende, en la calidad del aire que respira la población del AMLC (Pacsi 2016, Silva *et al.* 2017).



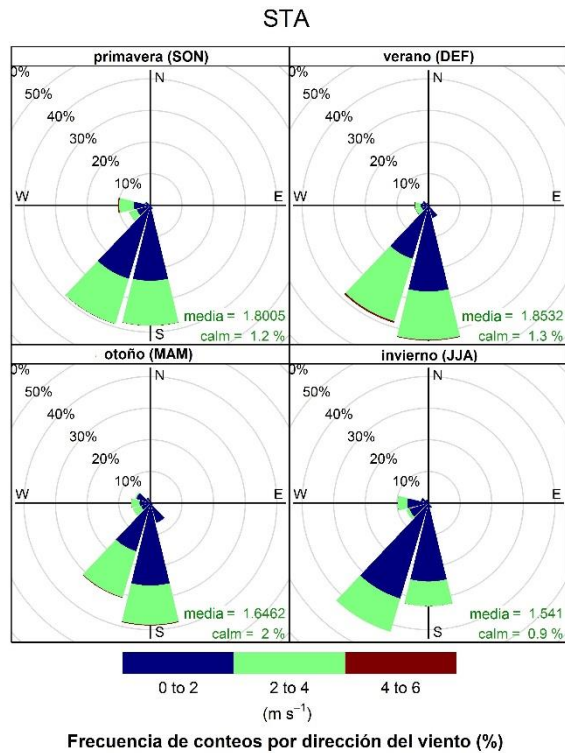
**Figura 16: Rosa de vientos para la estación SBJ, periodo 2015-2019**

Las rosas de vientos para la estación CDM se presentan en la **Figura 17**. La dirección predominante del viento en las cuatro estaciones del año es la del suroeste (SO), variando su velocidad en promedio de 2.55, 2.39, 2.03 y 1.93 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha evidenciado que las velocidades de viento en las estaciones primavera y verano son mayores en comparación a las estaciones de otoño e invierno. Estas diferencias influyen en la dispersión de los contaminantes atmosféricos y, por ende, en la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).

Las rosas de vientos para la estación STA se presentan en la **Figura 18**. La dirección predominante del viento en las cuatro estaciones del año son la del sur (S) y suroeste (SO), variando su velocidad en promedio de 1.80, 1.85, 1.65 y 1.54 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha evidenciado que, las velocidades de viento en las estaciones primavera y verano son mayores en comparación a las estaciones de otoño e invierno. Lo cual guarda relación con la intensificación del Anticiclón del Pacífico Sur y su consecuente estabilidad atmosférica que genera (Silva *et al.* 2018). Estas características influyen en el transporte de los contaminantes atmosféricos y, por ende, en la calidad del aire que respira la población del AMLC (Pacsi 2016, Silva *et al.* 2017).



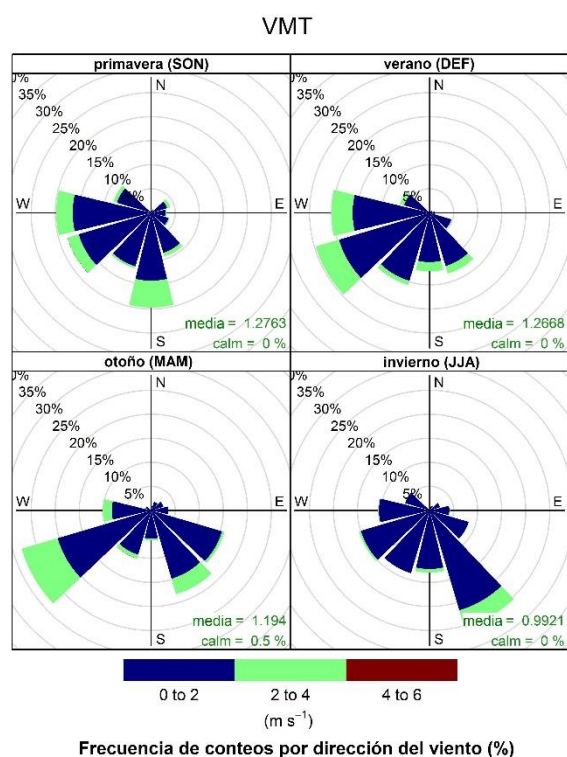
**Figura 17: Rosa de vientos para la estación CDM, periodo 2015-2019**



**Figura 18: Rosa de vientos para la estación STA, periodo 2015-2019**

Las rosas de vientos para la estación VMT se presentan en la **Figura 19**. Presenta una mayor variabilidad en la dirección del viento; sin embargo, la dirección predominante del viento durante el periodo de estudio, son la del sur (S) y suroeste (SO), variando su velocidad en

promedio de 1.28, 1.27, 1.19 y 0.99 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha observado mayores valores, en las estaciones de primavera y verano, en comparación con el otoño e invierno. Lo cual guarda relación con la acentuación del Anticiclón del Pacífico Sur y su consecuente inversión de subsidencia que genera (Silva *et al.* 2018). Estas características influyen en la dispersión de los contaminantes atmosféricos y, por ende, en la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).



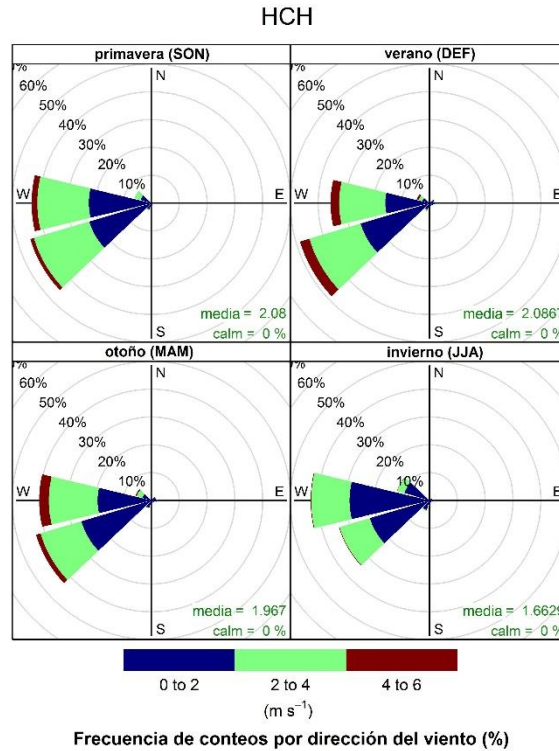
**Figura 19: Rosa de vientos para la estación VMT, periodo 2015-2019**

Las rosas de vientos para la estación HCH se presentan en la **Figura 20**. La dirección predominante del viento en las cuatro estaciones del año son la del suroeste (SO) y oeste (O), variando su velocidad en promedio de 2.08, 2.09, 1.97 y 1.66 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha evidenciado que, las velocidades de viento en las estaciones primavera y verano son mayores que, en las estaciones de otoño e invierno. Lo cual guarda relación con la intensificación del Anticiclón del Pacífico Sur y su consecuente estabilidad atmosférica que genera (Silva *et al.* 2018). Estas características influyen en el transporte de los contaminantes atmosféricos y, por ende, en la calidad del aire que respira la población del AMLC (Pacsi 2016, Silva *et al.* 2017).

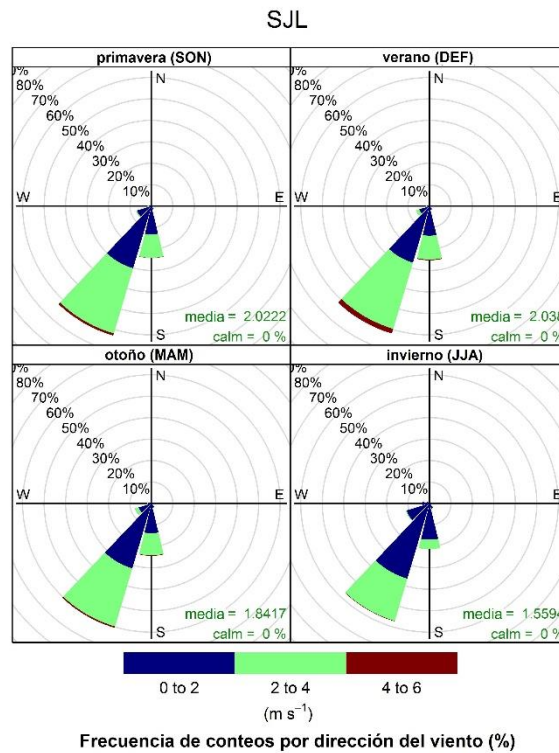
Las rosas de vientos para la estación SJL se presentan en la **Figura 21**. La dirección predominante del viento en las cuatro estaciones del año es la del suroeste (SO), variando su velocidad en promedio de 2.02, 2.04, 1.84 y 1.56 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se observa mayores valores en las estaciones de



primavera y verano, en comparación con el otoño e invierno. Lo cual guarda relación con la acentuación del Anticiclón del Pacífico Sur y su consecuente inversión de subsidencia que genera (Silva *et al.* 2018). Estas características influyen en la dispersión de los contaminantes atmosféricos y, por ende, en la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).

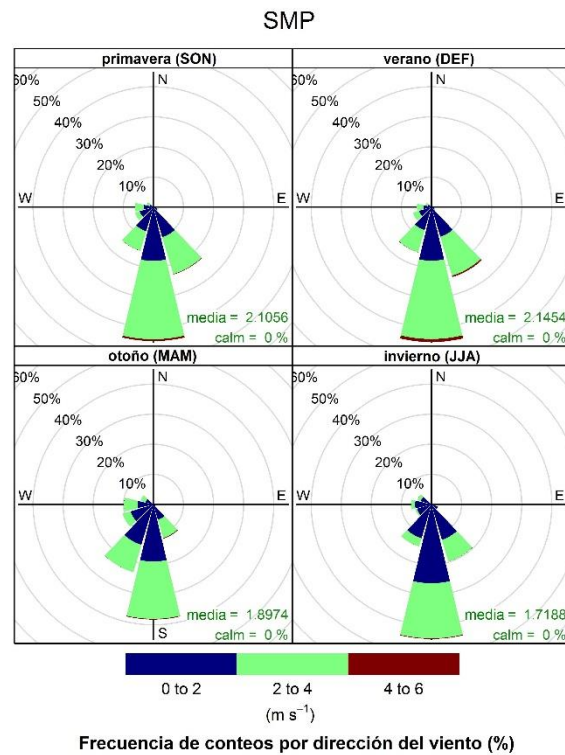


**Figura 20: Rosa de vientos para la estación HCH, periodo 2015-2019**



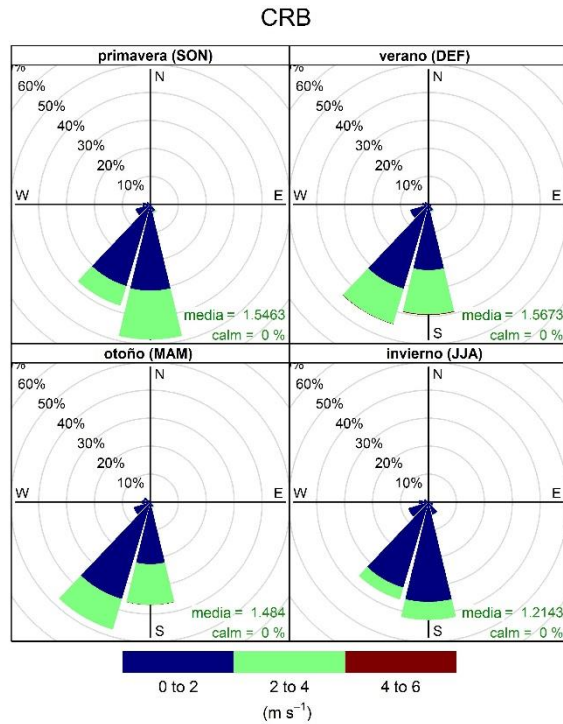
**Figura 21: Rosa de vientos para la estación SJL, periodo 2015-2019**

Las rosas de vientos para la estación SMP se presentan en la **Figura 22**. La dirección predominante del viento en las cuatro estaciones del año es la del sur (S), variando su velocidad en promedio de 2.10, 2.14, 1.89 y 1.72 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha evidenciado que, las velocidades del viento en las estaciones primavera y verano son mayores en comparación a las estaciones de otoño e invierno. Lo cual guarda relación con la intensificación del Anticiclón del Pacífico Sur y su consecuente estabilidad atmosférica que genera (Silva *et al.* 2018). Estas características influyen en el transporte de los contaminantes atmosféricos y, por ende, en la calidad del aire que respira la población del AMLC (Pacsi 2016, Silva *et al.* 2017).

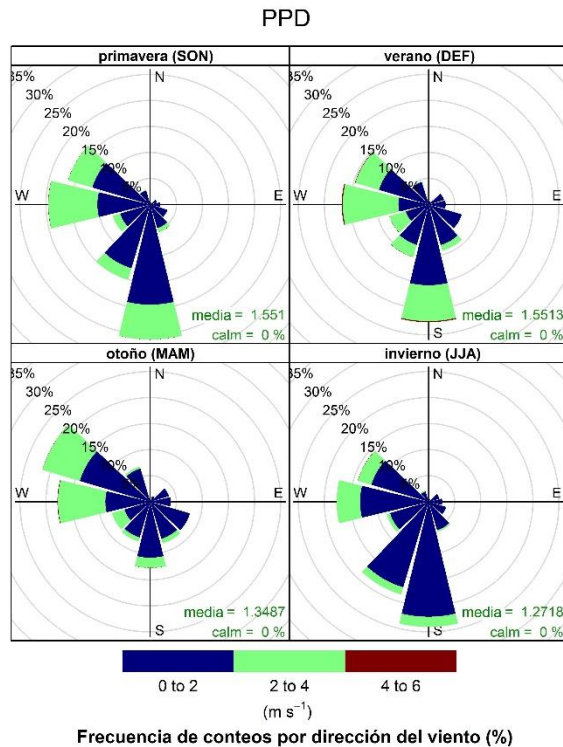


**Figura 22: Rosa de vientos para la estación SMP, periodo 2015-2019**

Las rosas de vientos para la estación CRB se presentan en la **Figura 23**. La dirección predominante del viento en las cuatro estaciones del año son la del sur (S) y suroeste (SO), variando su velocidad en promedio de 1.55, 1.57, 1.48 y 1.21 m/s en las estaciones primavera, verano, otoño e invierno, respectivamente. Se ha evidenciado que, las velocidades de viento en las estaciones primavera y verano son mayores en comparación a las estaciones de otoño e invierno. Lo cual guarda relación con la acentuación del Anticiclón del Pacífico Sur y su consecuente inversión de subsidencia que genera (Silva *et al.* 2018). Estas características influyen en la dispersión de los contaminantes atmosféricos y, por ende, en la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).



**Figura 23: Rosa de vientos para la estación CRB, periodo 2015-2019**



**Figura 24: Rosa de vientos para la estación PPD periodo 2015-2019**

Las rosas de vientos para la estación PPD se presentan en la **Figura 24**. Se evidencia alta variabilidad en la dirección del viento; sin embargo, la dirección predominante en las cuatro estaciones del año son la del sur (S) y sureste (SO), variando su velocidad en promedio de 1.55, 1.55, 1.35 y 1.27 m/s en las estaciones primavera, verano, otoño e invierno,

respectivamente. Se ha evidenciado que, las velocidades de viento en las estaciones primavera y verano son mayores en comparación a las estaciones de otoño e invierno. Lo cual guarda relación con la intensificación del Anticiclón del Pacífico Sur y su consecuente estabilidad atmosférica que genera (Silva *et al.* 2018). Estas características influyen en el transporte de los contaminantes atmosféricos y, por ende, en la calidad del aire que respira la población del AMLC (Pacsi 2016, Silva *et al.* 2017).

#### 4.2.1 Patrones espaciales de las variables meteorológicas identificados a través del análisis de componentes principales (ACP)

Con el fin de aplicar el ACP en la determinación de los patrones espaciales de las variables meteorológicas, primero se realizó el análisis de las correlaciones entre las estaciones meteorológicas por cada una de las variables, para luego determinar las pruebas de KMO y esfericidad de Bartlett.

En la **Tabla 29** se presentan los resultados de la correlación de Pearson para los datos de temperatura del aire entre las estaciones de monitoreo administradas por el SENAMHI en el AMLC. En general, se observa altos valores de correlación entre todas las estaciones. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de HCH - SJL y HCH - ATE, ambos con un valor de 0.99. Las dos correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este del AMLC, por lo que comparten características meteorológicas similares respecto a la temperatura del aire (Pacsi 2016, Espinoza 2018).

**Tabla 29: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de temperatura del aire**

Estaciones	ATE	STA	SMP	SJL	SBJ	PPD	HCH	VMT	CRB	CDM
ATE	1									
STA	<b>0.98</b>	1								
SMP	<b>0.63</b>	<b>0.65</b>	1							
SJL	<b>0.99</b>	<b>0.98</b>	<b>0.65</b>	1						
SBJ	<b>0.95</b>	<b>0.96</b>	<b>0.74</b>	<b>0.95</b>	1					
PPD	<b>0.97</b>	<b>0.97</b>	<b>0.63</b>	<b>0.98</b>	<b>0.94</b>	1				
HCH	<b>0.99</b>	<b>0.99</b>	<b>0.66</b>	<b>0.99</b>	<b>0.96</b>	<b>0.97</b>	1			
VMT	<b>0.80</b>	<b>0.81</b>	<b>0.15</b>	<b>0.80</b>	<b>0.74</b>	<b>0.80</b>	<b>0.79</b>	1		
CRB	<b>0.98</b>	<b>0.98</b>	<b>0.65</b>	<b>0.98</b>	<b>0.95</b>	<b>0.99</b>	<b>0.98</b>	<b>0.79</b>	1	
CDM	<b>0.92</b>	<b>0.93</b>	<b>0.70</b>	<b>0.92</b>	<b>0.95</b>	<b>0.91</b>	<b>0.92</b>	<b>0.78</b>	<b>0.93</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 30** se presentan los resultados de la correlación de Pearson para los datos de humedad relativa entre las estaciones de monitoreo administradas por el SENAMHI en el AMLC. Se observan altos valores de correlación entre todas las estaciones. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de HCH - ATE y SJL - STA, con valores de 0.98 y 0.97, respectivamente. Ambas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este del AMLC, por lo que presentan comportamientos meteorológicos similares respecto la humedad relativa (Pacsi 2016, Silva *et al.* 2017).

**Tabla 30: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de humedad relativa**

Estaciones	ATE	STA	SMP	SJL	SBJ	PPD	HCH	CRB	CDM	VMT
ATE	1									
STA	<b>0.92</b>	1								
SMP	<b>0.86</b>	<b>0.81</b>	1							
SJL	<b>0.92</b>	<b>0.97</b>	<b>0.82</b>	1						
SBJ	<b>0.90</b>	<b>0.86</b>	<b>0.94</b>	<b>0.85</b>	1					
PPD	<b>0.94</b>	<b>0.91</b>	<b>0.88</b>	<b>0.92</b>	<b>0.91</b>	1				
HCH	<b>0.98</b>	<b>0.92</b>	<b>0.87</b>	<b>0.92</b>	<b>0.92</b>	<b>0.95</b>	1			
CRB	<b>0.95</b>	<b>0.90</b>	<b>0.88</b>	<b>0.91</b>	<b>0.91</b>	<b>0.97</b>	<b>0.96</b>	1		
CDM	<b>0.84</b>	<b>0.80</b>	<b>0.96</b>	<b>0.81</b>	<b>0.93</b>	<b>0.85</b>	<b>0.85</b>	<b>0.86</b>	1	
VMT	<b>0.89</b>	<b>0.90</b>	<b>0.85</b>	<b>0.89</b>	<b>0.89</b>	<b>0.86</b>	<b>0.89</b>	<b>0.85</b>	<b>0.86</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 31** se presentan los resultados de la correlación de Pearson para los datos de dirección del viento entre las estaciones de monitoreo administradas por el SENAMHI en el AMLC. En general, se observan valores bajos de correlación entre todas las estaciones. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de SBJ - SMP y VMT - SMP, con valores de 0.58 y 0.47, respectivamente. Ambas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Las estaciones SMP, SBJ y VMT, se encuentran en las zonas norte, centro y sur del AMLC, respectivamente (Pacsi 2016).

En la **Tabla 32** se muestran los resultados de la correlación de Pearson para los datos de velocidad del viento entre las estaciones de monitoreo administradas por el SENAMHI en el AMLC. En general, se observan valores intermedios de correlación entre todas las

estaciones. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de CRB - SJL y HCH - ATE, con valores de 0.79 y 0.75, respectivamente. Ambas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona este y noreste del AMLC, por lo que comparten características meteorológicas similares respecto a la velocidad del viento (Pacsi 2016, Sánchez y Ordoñez 2016).

**Tabla 31: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de dirección del viento**

Estaciones	ATE	STA	SMP	SJL	SBJ	PPD	HCH	CRB	CDM	VMT
ATE	1									
STA	<b>0.11</b>	1								
SMP	-0.02	<b>0.42</b>	1							
SJL	<b>0.09</b>	<b>0.12</b>	<b>0.19</b>	1						
SBJ	0.03	<b>0.35</b>	<b>0.58</b>	<b>0.18</b>	1					
PPD	<b>0.13</b>	<b>0.20</b>	<b>0.39</b>	<b>0.24</b>	<b>0.33</b>	1				
HCH	<b>0.35</b>	<b>0.07</b>	<b>0.09</b>	<b>0.19</b>	<b>0.18</b>	<b>0.26</b>	1			
CRB	0.00	0.01	0.04	-0.01	-0.02	-0.03	<b>-0.10</b>	1		
CDM	-0.01	0.01	<b>0.11</b>	<b>0.06</b>	<b>0.19</b>	0.03	<b>0.07</b>	-0.01	1	
VMT	0.01	<b>0.29</b>	<b>0.47</b>	<b>0.11</b>	<b>0.44</b>	<b>0.23</b>	<b>0.12</b>	0.05	<b>-0.07</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

**Tabla 32: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de velocidad del viento**

Estaciones	ATE	STA	SMP	SJL	SBJ	PPD	HCH	CDM	VMT	CRB
ATE	1									
STA	<b>0.61</b>	1								
SMP	<b>0.48</b>	<b>0.63</b>	1							
SJL	<b>0.68</b>	<b>0.73</b>	<b>0.66</b>	1						
SBJ	<b>0.28</b>	<b>0.54</b>	<b>0.70</b>	<b>0.47</b>	1					
PPD	<b>0.53</b>	<b>0.56</b>	<b>0.45</b>	<b>0.55</b>	<b>0.30</b>	1				
HCH	<b>0.75</b>	<b>0.73</b>	<b>0.48</b>	<b>0.70</b>	<b>0.33</b>	<b>0.59</b>	1			
CDM	<b>0.58</b>	<b>0.57</b>	<b>0.62</b>	<b>0.63</b>	<b>0.41</b>	<b>0.48</b>	<b>0.59</b>	1		
VMT	<b>0.42</b>	<b>0.55</b>	<b>0.60</b>	<b>0.53</b>	<b>0.63</b>	<b>0.41</b>	<b>0.45</b>	<b>0.48</b>	1	
CRB	<b>0.67</b>	<b>0.67</b>	<b>0.61</b>	<b>0.79</b>	<b>0.38</b>	<b>0.59</b>	<b>0.69</b>	<b>0.66</b>	<b>0.48</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

En la **Tabla 33** se presentan los resultados de la correlación de Pearson para los datos de radiación solar entre las estaciones de monitoreo administradas por el SENAMHI en el AMLC. Se observan valores altos de correlación entre todas las estaciones. Los mayores coeficientes de correlación de Pearson se presentan entre las estaciones de CRB - PPD y HCH - ATE, con valores de 0.98 y 0.97, respectivamente. Ambas correlaciones son estadísticamente significativas con un nivel de significancia de 5%. Estas estaciones se encuentran ubicadas en la zona norte y este del AMLC, respectivamente, por lo que presentan comportamientos meteorológicos similares respecto la radiación solar (Sánchez *et al.* 2015, Pacsi 2016).

**Tabla 33: Matriz de correlación (r de Pearson) entre las estaciones de monitoreo para los datos de radiación solar**

Estaciones	ATE	SMP	SJL	PPD	HCH	CRB	VMT
ATE	1						
SMP	<b>0.90</b>	1					
SJL	<b>0.96</b>	<b>0.93</b>	1				
PPD	<b>0.93</b>	<b>0.94</b>	<b>0.95</b>	1			
HCH	<b>0.97</b>	<b>0.92</b>	<b>0.96</b>	<b>0.95</b>	1		
CRB	<b>0.93</b>	<b>0.94</b>	<b>0.95</b>	<b>0.98</b>	<b>0.94</b>	1	
VMT	<b>0.86</b>	<b>0.87</b>	<b>0.86</b>	<b>0.87</b>	<b>0.87</b>	<b>0.88</b>	1

Los valores en negrita corresponden a coeficientes de correlación estadísticamente significativos

Finalmente, en la **Tabla 34** se muestran los resultados de las pruebas de KMO y esfericidad de Bartlett aplicado sobre los datos de variables meteorológicas entre las estaciones de monitoreo en el AMLC, administradas por el SENAMHI. Se puede observar que, las medidas Kaiser-Meyer-Olkin para las cinco variables meteorológicas son mayores a 0.5, lo que nos indica la idoneidad de los datos para el análisis estadístico multivariado (Abdullah *et al.* 2018). Así también, los niveles de significancia de la prueba de esfericidad de Bartlett son menores al 5%, lo que confirma que, las estaciones de monitoreo están correlacionadas entre sí y, por lo tanto, es factible realizar el análisis de componentes principales (ACP), sobre las variables meteorológicas entre las estaciones de monitoreo (Nazmi *et al.* 2020).

**Tabla 34: Pruebas de KMO y esfericidad de Bartlett de las variables meteorológicas**

Prueba de KMO y Bartlett	Temperatura del aire	Humedad relativa	Dirección del viento	Velocidad del viento	Radiación solar	
Medida Kaiser-Meyer- Olkin	0.905	0.933	0.737	0.921	0.936	
Prueba de esfericidad de Bartlett	Aprox. Chi- cuadrado gl Sig.	78402.8 45 0.000	65087.9 45 0.000	1386.1 45 0.000	6390.7 45 0.000	40781.9 21 0.000

La **Tabla 35** muestra los principales resultados de la aplicación del ACP para los datos temperatura del aire en todas las estaciones de monitoreo, durante el periodo 2015-2019. Considerando valores propios mayores a 1 (criterio de Kaiser), solo se seleccionó una CP, que explica el 88.3% de la varianza de datos original (Zhao *et al.* 2015). Así también, se extrajo un segundo CP, el cual captura el 97% de la varianza acumulada (Castro y Pires 2019). Las cargas de factores rotados (logrados por el algoritmo de rotación *varimax*) remarcados en negrita indican la influencia de cada estación en las CP.

**Tabla 35: Principales resultados de la aplicación del ACP a los datos de temperatura del aire**

	CP1	CP2
ATE	<b>0.86</b>	0.49
STA	<b>0.86</b>	0.50
SMP	0.19	<b>0.98</b>
SJL	<b>0.85</b>	0.50
SBJ	0.77	0.61
PPD	<b>0.85</b>	0.48
HCH	<b>0.84</b>	0.52
VMT	<b>0.98</b>	-0.05
CRB	0.85	0.51
CDM	0.78	0.55
Autovalores	8.83	0.87
% de varianza	88.3	8.7
% acumulado	88.3	97.0

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente



Las cargas factoriales representan la ponderación de las variables originales para cada CP, así como la correlación entre las variables originales y las CP. La rotación *varimax* fue el algoritmo de rotación utilizado para maximizar las cargas cuadradas de las columnas, lo cual permitió obtener las cargas factoriales rotadas. Por lo tanto, cada CP presenta altas cargas para algunas variables y bajas cargas para otras, lo que facilita la interpretación de éstas en términos de las variables originales (Pires *et al.* 2009).

Las cargas factoriales rotadas marcadas en negrita indican las variables originales que más influyen en el CP correspondiente. Por lo tanto, la CP1 tuvo contribuciones importantes de las estaciones VMT, STA, ATE, PPD, SJL, CRB, HCH, CDM y SBJ; mientras que, la CP2 estuvo muy cargada por la estación SMP. Los sitios de monitoreo agrupados en la misma CP presentaron patrones espaciales similares respecto al parámetro meteorológico temperatura del aire (Pires *et al.* 2009).

La **Tabla 36** muestra los principales resultados de la aplicación del ACP para los datos de humedad relativa en todas las estaciones de monitoreo, durante el periodo 2015-2019. La primera CP capturó el 90.5% de la variabilidad de los datos, así también posee un valor propio de 9.05, cumpliendo de esta manera con el criterio de Kaiser y el de la varianza acumulada superior al 85%. Para realizar el ACJ se decidió elegir adicionalmente una segunda CP, con esto se logró una varianza acumulada del 94.6% de los datos de humedad relativa. La alta variabilidad que capturan pocas CP, guarda relación con los altos valores de correlación presentados en la **Tabla 30** (Wang *et al.* 2018).

Las cargas factoriales rotadas marcadas en negrita indican las variables originales que más influyen en la CP correspondiente. Por lo tanto, la CP1 tuvo contribuciones importantes de las estaciones SJL, STA, ATE, HCH, PPD, CRB y VMT; mientras que, la CP2 estuvo muy cargada por las estaciones CDM, SMP y SBJ. Los sitios de monitoreo agrupados en la misma CP presentaron patrones espaciales similares respecto al parámetro meteorológico humedad relativa (Castro y Pires 2019).

La **Tabla 37** presenta los principales resultados de la aplicación del ACP para los datos de dirección del viento en todas las estaciones de monitoreo, durante el periodo 2015-2019. Considerando valores propios mayores a uno (criterio de Kaiser), solo se seleccionaron tres CP, las cuales explican el 52.0% de la varianza de los datos originales. Con el objetivo de cumplir el segundo criterio (varianza acumulada superior al 85%) se seleccionaron

adicionalmente cuatro CP, con esto se logró una varianza acumulada del 85.8% de los datos originales de dirección del viento. La baja variabilidad que capturan las CP, guarda relación con los bajos valores de correlación presentados en la **Tabla 31** (Lu *et al.* 2011).

**Tabla 36: Principales resultados de la aplicación del ACP a los datos de humedad relativa**

	CP1	CP2
ATE	<b>0.81</b>	0.54
STA	<b>0.87</b>	0.45
SMP	0.49	<b>0.85</b>
SJL	<b>0.87</b>	0.45
SBJ	0.59	<b>0.78</b>
PPD	<b>0.78</b>	0.57
HCH	<b>0.80</b>	0.56
CRB	<b>0.77</b>	0.58
CDM	0.46	<b>0.87</b>
VMT	<b>0.70</b>	0.62
Autovalores	9.05	0.41
% de varianza	90.5	4.1
% acumulado	90.5	94.6

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

**Tabla 37: Principales resultados de la aplicación del ACP a los datos de dirección del viento**

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
ATE	-0.13	<b>0.85</b>	0.01	-0.04	0.25	0.08	0.00
STA	0.27	0.06	0.07	-0.01	<b>0.89</b>	-0.02	0.06
SMP	<b>0.67</b>	-0.09	0.36	0.13	0.34	0.04	0.09
SJL	0.09	0.08	0.10	0.03	0.05	0.00	<b>0.98</b>
SBJ	<b>0.73</b>	0.06	0.22	0.28	0.20	-0.06	0.07
PPD	0.19	0.14	<b>0.94</b>	-0.02	0.06	-0.01	0.10
HCH	0.22	<b>0.77</b>	0.16	0.08	-0.22	-0.14	0.13
CRB	0.03	-0.03	-0.01	0.00	-0.01	<b>0.99</b>	0.00
CDM	0.03	0.02	-0.01	<b>0.97</b>	-0.01	0.00	0.03
VMT	<b>0.88</b>	0.06	-0.04	-0.21	0.03	0.05	0.03
Autovalores	2.71	1.40	1.08	0.99	0.90	0.79	0.70
% de varianza	27.1	14.0	10.8	9.9	9.0	7.9	7.0
% acumulado	27.1	41.1	52.0	61.9	70.9	78.7	85.8

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Las cargas factoriales rotadas en negrita indican las variables originales que más influyen en el CP correspondiente. Por lo tanto, la CP1 tuvo contribuciones importantes de las estaciones VMT, SBJ y SMP; mientras que, la CP2 estuvo muy cargada por las estaciones ATE y HCH. Las estaciones PPD, CDM, STA, CRB y SJL se asociaron significativamente con las CP3, CP4, CP5, CP6 y CP7, respectivamente. Los sitios de monitoreo agrupados en la misma CP presentaron patrones espaciales similares respecto el parámetro meteorológico dirección del viento (Pires *et al.* 2009).

La **Tabla 38** muestra los principales resultados de la aplicación del ACP para los datos de velocidad del viento en todas las estaciones de monitoreo, durante el periodo 2015-2019. Considerando valores propios mayores a uno (criterio de Kaiser), solo se seleccionaron dos CP, las cuales explican el 72.7% de la varianza de los datos originales. Con el objetivo de cumplir el segundo criterio (varianza acumulada superior al 85%) se seleccionaron adicionalmente tres CP, con esto se logró una varianza acumulada del 87.3% de los datos originales de velocidad del viento. La baja variabilidad que capturan las CP, guarda relación con los valores intermedios de correlación presentados en la **Tabla 32** (Wang *et al.* 2018).

**Tabla 38: Principales resultados de la aplicación del ACP a los datos de velocidad del viento**

	CP1	CP2	CP3	CP4	CP5
ATE	<b>0.82</b>	0.04	0.29	0.13	0.22
STA	<b>0.70</b>	0.52	0.07	0.25	0.10
SMP	0.27	<b>0.74</b>	0.44	0.16	0.12
SJL	<b>0.70</b>	0.43	0.34	0.20	0.03
SBJ	0.12	<b>0.88</b>	0.08	0.07	0.29
PPD	0.36	0.15	0.18	<b>0.89</b>	0.13
HCH	<b>0.84</b>	0.13	0.17	0.26	0.19
CDM	0.38	0.22	<b>0.81</b>	0.16	0.19
VMT	0.25	0.45	0.18	0.15	<b>0.79</b>
CRB	<b>0.64</b>	0.30	0.46	0.30	-0.01
Autovalores	6.10	1.16	0.56	0.48	0.43
% de varianza	61.0	11.6	5.6	4.8	4.3
% acumulado	61.0	72.7	78.2	83.0	87.3

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Las cargas factoriales rotadas marcadas en negrita indican las variables originales que más influyen en el CP correspondiente. Por lo tanto, la CP1 tuvo contribuciones importantes de las estaciones HCH, ATE, STA, SJL y CRB; mientras que, la CP2 estuvo muy cargada por

las estaciones SBJ y SMP. Las estaciones CDM, PPD y VMT se asociaron significativamente con las CP3, CP4 y CP5, respectivamente. Los sitios de monitoreo agrupados en la misma CP presentaron patrones espaciales similares respecto al parámetro meteorológico velocidad del viento (Pires *et al.*, 2009).

La **Tabla 39** muestra los principales resultados de la aplicación del ACP para los datos de radiación solar en todas las estaciones de monitoreo disponibles, durante el periodo 2015-2019. La primera CP capturó el 93.4% de la variabilidad de los datos, el cual posee un valor propio de 6.54, cumpliendo de esta manera con el criterio de Kaiser y el de la varianza acumulada superior al 85%. Para realizar el ACJ se decidió elegir adicionalmente una segunda CP, con esto se logró una varianza acumulada del 96.1% de los datos de humedad relativa. La alta variabilidad que capturan pocas CP, guarda relación con los altos valores de correlación presentados en la **Tabla 33** (Zhao *et al.* 2015).

**Tabla 39: Principales resultados de la aplicación del ACP a los datos de radiación solar**

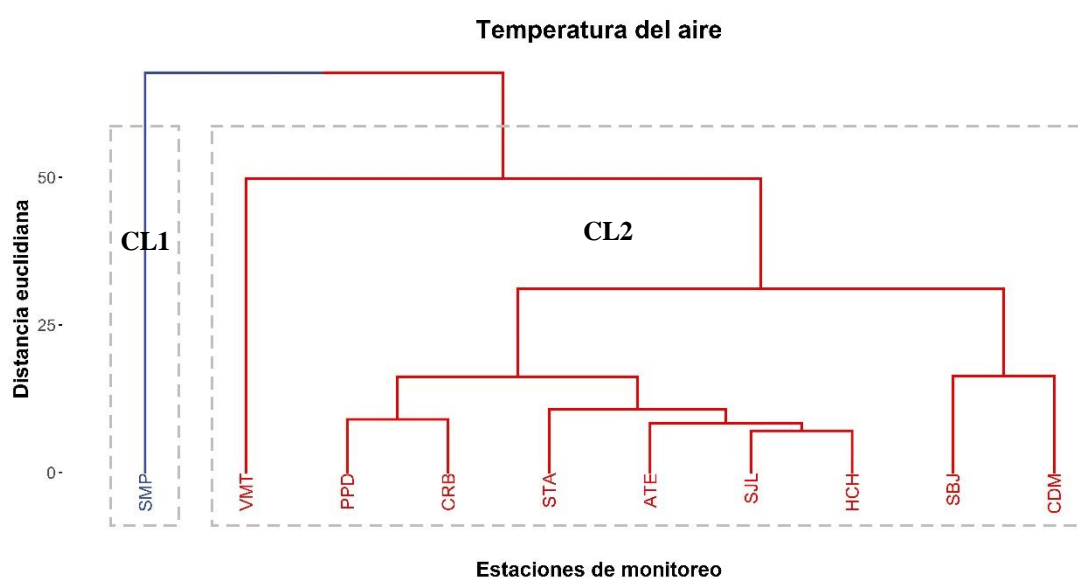
	CP1	CP2
ATE	<b>0.84</b>	0.49
SMP	<b>0.75</b>	0.60
SJL	<b>0.85</b>	0.49
PPD	<b>0.82</b>	0.54
HCH	<b>0.84</b>	0.51
CRB	<b>0.81</b>	0.55
VMT	0.51	<b>0.86</b>
Autovalores	6.54	0.19
% de varianza	93.4	2.7
% acumulado	93.4	96.1

Los valores en negrita indican las estaciones que más influyen en el componente principal correspondiente

Las cargas factoriales rotadas marcadas en negrita indican las variables originales que más influyen en el CP correspondiente. Por lo tanto, la CP1 tuvo contribuciones importantes de las estaciones SJL, ATE, HCH, PPD, CRB y SMP; mientras que, la CP2 estuvo muy cargada por la estación VMT. Los sitios de monitoreo agrupados en la misma CP presentaron patrones espaciales similares respecto al parámetro meteorológico radiación solar (Abdullah *et al.* 2018, Castro y Pires 2019).

#### 4.2.2 Patrones espaciales de las variables meteorológicas identificados a través del análisis de clúster jerárquico (ACJ)

La **Figura 25** muestra los resultados del método del ACJ sobre los datos de temperatura del aire. De acuerdo con el valor de la distancia euclidiana de la temperatura del aire calculado a partir de software R, las diez estaciones de la REMCA pueden ser agrupados en dos grupos: clúster I (CL1)-estación SMP y clúster II (CL2)-estaciones VMT, PPD, CRB, STA, ATE, SJL, HCH, SBJ y CDM. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 35**), donde los componentes principales CP1 y CP2, corresponden a los grupos CL2 y CL1 del ACJ, respectivamente.

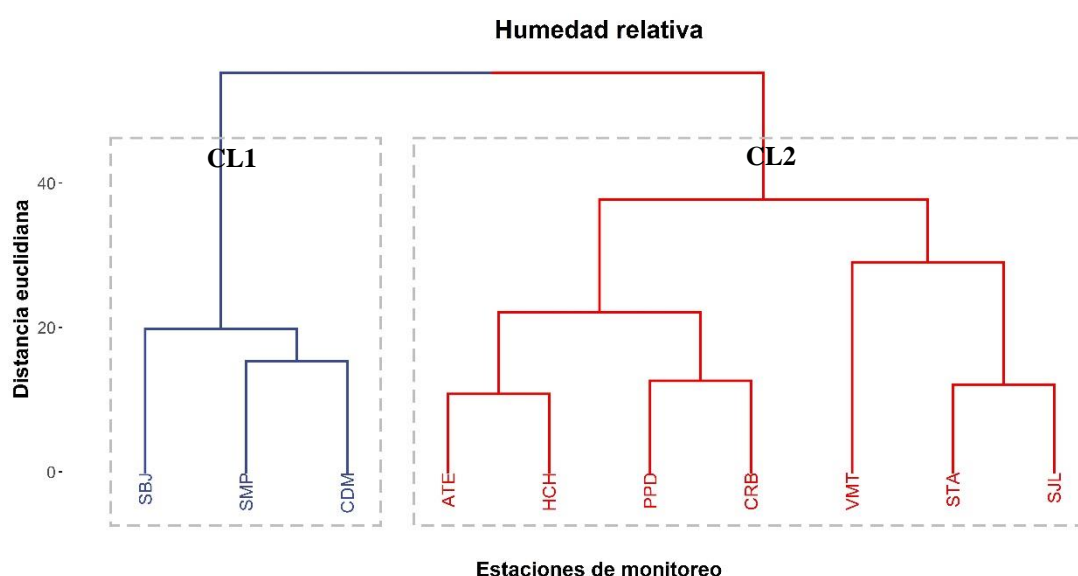


**Figura 25: Dendrograma resultante de la aplicación del ACJ a los datos de temperatura del aire**

Se puede encontrar que, estos grupos dependen evidentemente de la geografía y las condiciones meteorológicas particulares de las zonas. Las estaciones de monitoreo presentan alta homogeneidad entre ellas para la variable temperatura del aire, por lo que, los clústeres generados agrupan a dichas estaciones en grupos reducidos. Las estaciones que conforman el CL2 presentan mayor similitud respecto a su ubicación geográfica, por ejemplo, las estaciones de PPD y CRB, están ubicadas en la zona norte del AMLC; STA, ATE, SJL y HCH, se ubican en la zona este del AMLC y las estaciones SBJ y CDM, se encuentran ubicadas en la zona centro del AMLC; mientras que, la estación de SMP se agrupa individualmente en el CL1, presentando características diferentes al resto de las estaciones. Esto último, guarda relación con la existencia de microclimas en los diferentes distritos de

la ciudad, que influyen directamente en las condiciones de dispersión de los contaminantes atmosféricos y el nivel de la calidad del aire en el AMLC (Ibáñez 2008, Pacsi 2016).

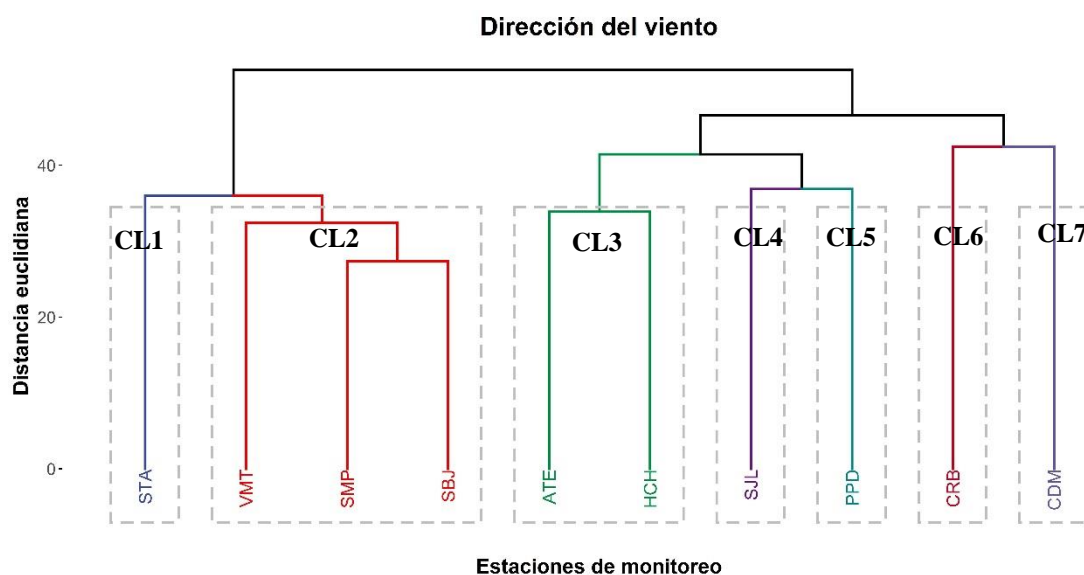
La **Figura 26** muestra los resultados del método del ACJ sobre los datos de humedad relativa. De acuerdo con el valor de la distancia euclidiana de la humedad relativa calculado a partir de software R, las diez estaciones de la REMCA pueden ser agrupados en dos grupos: clúster I (CL1)-estaciones SBJ, SMP y CDM; y clúster II (CL2)-estaciones ATE, HCH, PPD, CRB, STA, SJL y ATE. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 36**), donde los componentes principales CP1 y CP2, correspondían a los grupos CL2 y CL1 del ACJ, respectivamente.



**Figura 26: Dendrograma resultante de la aplicación del ACJ a los datos de humedad relativa**

Se puede observar que, estos grupos dependen de la geografía y las condiciones meteorológicas particulares de las zonas de estudio. Las estaciones de monitoreo presentan alta homogeneidad entre ellas para la variable humedad relativa, por lo que, los clústeres generados agrupan a éstas en grupos reducidos. Las estaciones que conforman el CL2 presentan mayor similitud de acuerdo a su ubicación geográfica, por ejemplo, las estaciones ATE, HCH, STA y SJL se ubican en la zona este; mientras que, PPD y CRB, se ubican en la zona norte del AMLC. Por otro lado, las estaciones de SBJ, SMP y CDM se agrupan en el CL1, presentando características diferentes al resto de las estaciones. Esto último, guarda relación con la existencia de microclimas en los diferentes distritos de la ciudad, que influyen directamente en las condiciones de dispersión de los contaminantes atmosféricos y el nivel de la calidad del aire en el AMLC (Pacsi 2016, Espinoza 2018).

La **Figura 27** muestra los resultados del método del ACJ sobre los datos de dirección del viento. De acuerdo con el valor de la distancia euclidiana de la dirección del viento calculado a partir de software R, las diez estaciones de la REMCA pueden ser agrupados en siete grupos: clúster I (CL1)-estación STA; clúster II (CL2)-estaciones VMT, SMP y SBJ; clúster III (CL3)-estaciones ATE y HCH; clúster IV (CL4)-estación SJL; clúster V (CL5)-estación PPD; clúster VI (CL6)-estación CRB; y clúster VII (CL7)-estación CDM. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 37**), donde los componentes principales CP1, CP2, CP3, CP4, CP5, CP6 y CP7, correspondían a los grupos CL2, CL3, CL5, CL7, CL1, CL6 y CL4 del ACJ, respectivamente.

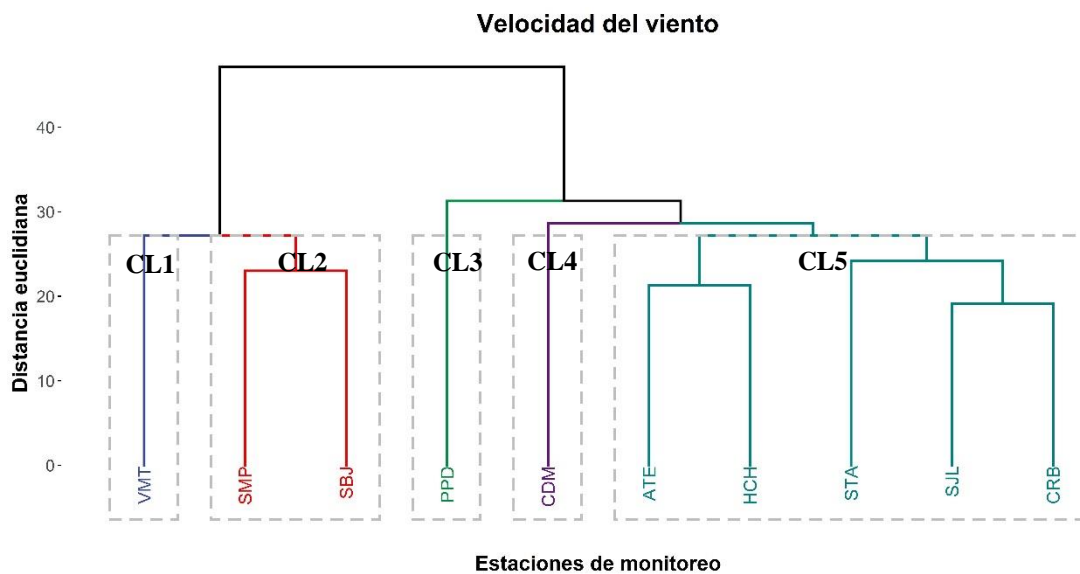


**Figura 27: Dendrograma resultante de la aplicación del ACJ a los datos de dirección del viento**

Se puede encontrar que, estos grupos obedecen a la geografía y las características meteorológicas de las zonas en las que están ubicadas. Las estaciones de monitoreo presentan alta heterogeneidad entre ellas para la variable dirección del viento, por lo que, los clústers generados agrupan a éstas en siete grupos. Las estaciones que conforman el CL2 (SMP, SBJ y VMT) se ubican en la zona norte, centro y sur del AMLC, respectivamente. Por otro lado, las estaciones ATE y HCH, se localizan en la zona este, y se agrupan en el CL3, presentando características diferentes al resto de las estaciones. Esto último, guarda relación con la existencia de microclimas en los diferentes distritos de la ciudad, que influyen directamente en las condiciones de dispersión de los contaminantes atmosféricos y el nivel de la calidad del aire en el AMLC (Ibáñez 2008, Pacsi 2016).

La **Figura 28** muestra los resultados del método del ACJ sobre los datos de velocidad del viento. De acuerdo con el valor de la distancia euclidiana de la velocidad del viento calculado

a partir de software R, las diez estaciones de la REMCA pueden ser agrupados en cinco grupos: clúster I (CL1)-estación VMT; clúster II (CL2)-estaciones SMP y SBJ; clúster III (CL3)-estación PPD; clúster IV (CL4)-estación CDM; y clúster V (CL5)-estación ATE, HCH, STA, SJL y CRB. Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (**Tabla 38**), donde los componentes principales CP1, CP2, CP3, CP4 y CP5, correspondían a los grupos CL5, CL2, CL4, CL3 y CL1 del ACJ, respectivamente.



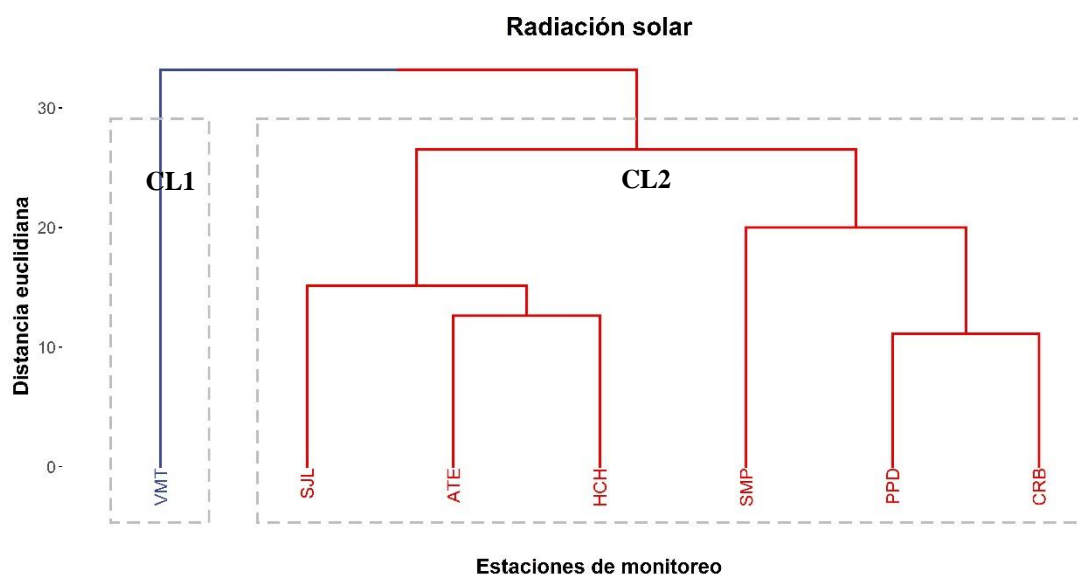
**Figura 28: Dendrograma resultante de la aplicación del ACJ a los datos de velocidad del viento**

Se puede observar que, estos grupos dependen de su ubicación geográfica y las condiciones meteorológicas particulares de las zonas donde se ubican. Las estaciones de monitoreo presentan moderada heterogeneidad entre ellas para la variable velocidad del viento, por lo que, los clústers generados agrupan a éstas en cinco grupos. Las estaciones que conforman el CL2 (SMP y SBJ) se ubican en la zona norte y centro del AMLC, respectivamente; mientras que, las estaciones de ATE, HCH, STA, SJL y CRB, se ubican en la zona este y norte, y se agrupan en el CL5, presentando características diferentes al resto de las estaciones. Esto último, guarda relación con la existencia de microclimas en los diferentes distritos de la ciudad, que influyen directamente en las condiciones de dispersión de los contaminantes atmosféricos y el nivel de la calidad del aire en el AMLC (Pacsi 2016, Silva *et al.* 2017).

La **Figura 29** muestra los resultados del método del ACJ sobre los datos de radiación solar. De acuerdo con el valor de la distancia euclidiana de la radiación solar calculado a partir de software R, las diez estaciones de la REMCA pueden ser agrupados en dos grupos: clúster I (CL1)-estación VMT y clúster II (CL2)-estaciones SJL, ATE, HCH, SMP, PPD y CRB.



Estos hallazgos coinciden con los resultados obtenidos por el método de ACP (Tabla 39), donde los componentes principales CP1 y CP2, correspondían a los grupos CL2 y CL1 del ACJ, respectivamente.



**Figura 29: Dendrograma resultante de la aplicación del ACJ a los datos de radiación solar**

Se evidencia que, estos grupos dependen de la geografía y las características meteorológicas de las zonas de las zonas de estudio. Las estaciones de monitoreo presentan alta homogeneidad entre ellas para la variable radiación solar, por lo que, los clústers generados agrupan a éstas en grupos reducidos. Las estaciones que conforman el CL2 presentan mayor similitud de acuerdo a su ubicación geográfica, por ejemplo, las estaciones SJL, ATE y HCH están ubicadas en la zona este; y las estaciones SMP, PPD y CRB, están ubicadas en la zona norte del AMLC. Por otro lado, la estación VMT se agrupa individualmente en el CL1, presentando características diferentes al resto de las estaciones. Esto último, guarda relación con la existencia de microclimas en los diferentes distritos de la ciudad, que influyen directamente en las condiciones de dispersión de los contaminantes atmosféricos y el nivel de la calidad del aire en el AMLC (Silva *et al.* 2017, 2018).

#### **4.3 PRONÓSTICO DE LAS CONCENTRACIONES DE MATERIAL PARTICULADO Y GASES EN EL ÁREA DE LIMA METROPOLITANA A TRAVÉS DE LA REGRESIÓN LINEAL MÚLTIPLE**

Para el pronóstico de las concentraciones horarias de PM<sub>10</sub>, PM<sub>2,5</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO y SO<sub>2</sub>, se eligió la estación ATE, debido a que esta estación es la que presenta los mayores niveles de

éstos contaminantes (**Tabla 14**), hallazgo también reportado por anteriores investigaciones (Pacsi 2016, Silva *et al.* 2017, 2018).

Los pronósticos generados se obtuvieron en base a dos modelos implementados para cada contaminante individualmente. El primer modelo se basó en la predicción del contaminante tomando como predictores a las concentraciones del mismo contaminante de las otras nueve estaciones de monitoreo (**Anexo 1**). El segundo modelo se implementó tomando como variables predictores a las variables meteorológicas, concentraciones de partículas y gases de la misma estación de calidad de aire (**Anexo 2**).

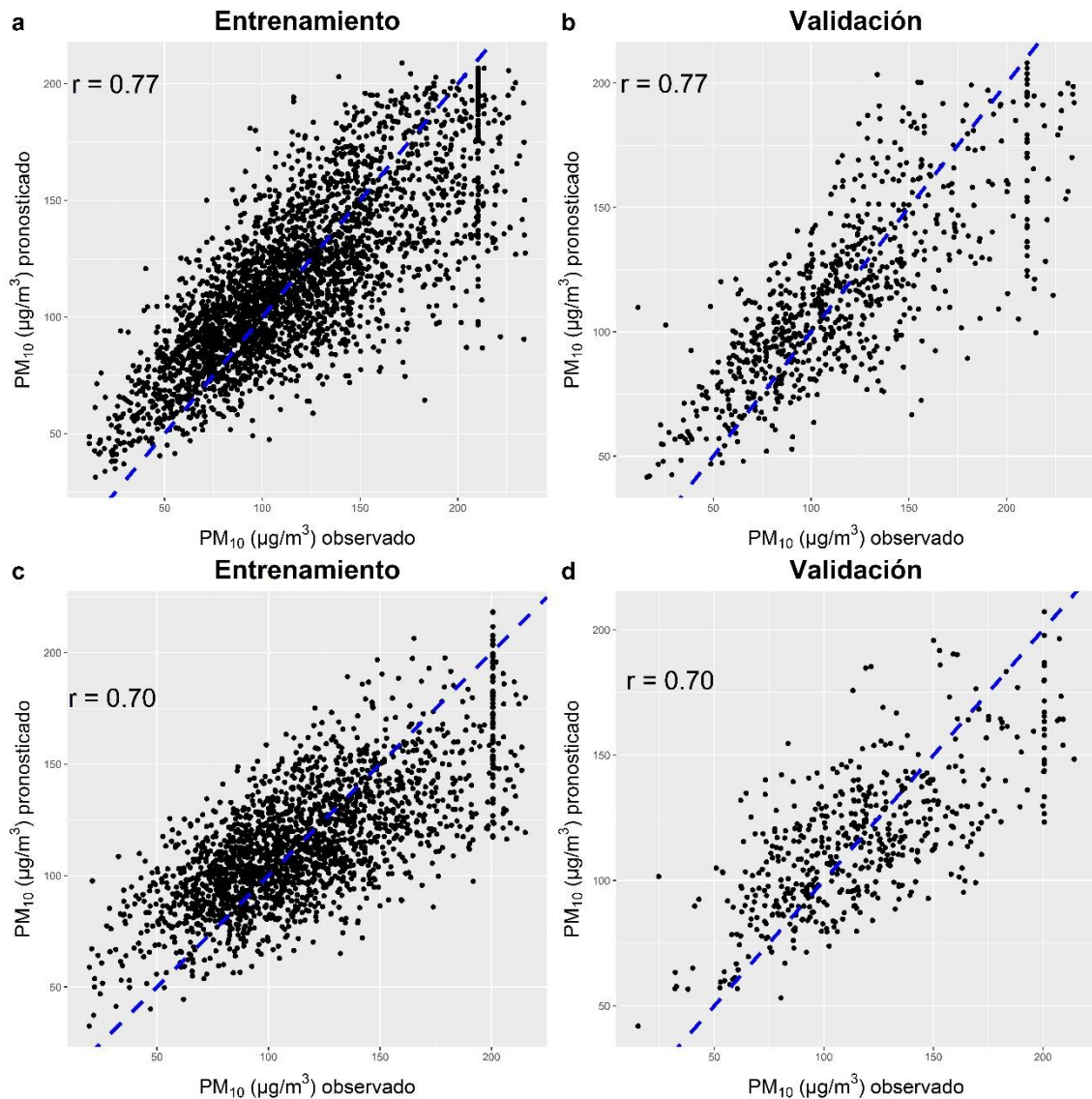
Para la implementación de todos los modelos generados se procedió a dividir el conjunto de datos en dos grupos, uno de entrenamiento (80 %) y el otro de validación (20 %). En todos los modelos implementados se consideraron únicamente como predictores aquellos que eran estadísticamente significativos al 5% (Abdullah *et al.* 2019). Esto dio como resultados modelos estadísticamente significativos con un nivel de confianza superior al 95%.

#### **4.3.1 Pronóstico del material particulado PM<sub>10</sub>**

En la **Figura 30** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de PM<sub>10</sub>, en la etapa de entrenamiento y validación. En la **Figura 30 a** y **b** se presenta el modelo de predicción para el PM<sub>10</sub>, implementado a partir de las concentraciones de PM<sub>10</sub> del resto de estaciones de calidad de aire. Los coeficientes de correlación de Pearson entre los valores observados y pronosticados, para ambas etapas son similares con un valor de 0.77. Asimismo, en la **Figura 30 c** y **d** se presenta el modelo de predicción para el PM<sub>10</sub>, implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. Los coeficientes de correlación de Pearson entre los valores observados y pronosticados, para ambas etapas son semejantes con un valor de 0.70. A pesar de tener valores de coeficientes de Pearson moderados, los modelos generados a partir de las variables meteorológicas no aportaron mayor captura de información, resultados similares se encontraron en (Ul-Saufie *et al.* 2013). Finalmente, la **Tabla 40** presenta los indicadores de desempeño de los modelos generados.

Para el primer modelo de pronóstico de PM<sub>10</sub> los predictores significativos fueron las concentraciones de PM<sub>10</sub> de las estaciones CDM (PM<sub>10</sub>CDM), STA(PM<sub>10</sub>STA), VMT(PM<sub>10</sub>VMT), HCH(PM<sub>10</sub>HCH) y SMP(PM<sub>10</sub>SMP). De la evaluación se determinó que, las concentraciones pronosticadas de PM<sub>10</sub> para la estación ATE aumentan en una relación de

0.776, 0.420 y 0.169 unidades cuando las variables  $PM_{10STA}$ ,  $PM_{10HCH}$  y  $PM_{10SMP}$  aumentan en una unidad, respectivamente; y disminuyen en 0.193 y 0.036 unidades por unidad de  $PM_{10CDM}$  y  $PM_{10VMT}$ , correspondientemente. Mientras tanto, el segundo modelo fue influenciado por nueve variables diferentes, es decir,  $PM_{2.5}$ , CO,  $NO_2$ ,  $O_3$ , T, HR, WS, WD y RS. Hubo un aumento de 0.813, 0.021, 0.476, 0.047 y una disminución de 1.405, 3.333, 1.294, 5.003, 0.053 unidades de  $PM_{10}$  pronosticado por unidad de  $PM_{2.5}$ , CO,  $NO_2$ , RS,  $O_3$ , T, HR, WS y WD, respectivamente (**Tabla 40**).



**Figura 30: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de  $PM_{10}$  para la estación ATE**

En la etapa de entrenamiento de los modelos, los coeficientes de determinación ( $r^2$ ) variaron de 0.60 a 0.49 para los modelos I y II, respectivamente. El menor valor de la raíz del error cuadrático medio (RECM), se presentó en el modelo II ( $28.2 \mu g/m^3$ ) respecto del modelo I ( $28.8 \mu g/m^3$ ). Mientras que, en la etapa de validación los coeficientes de determinación ( $r^2$ )

variaron de 0.59 a 0.48 para los modelos I y II, respectivamente; asimismo, el menor valor de la RECM se obtuvo en el modelo II ( $27.7 \mu\text{g}/\text{m}^3$ ) en comparación con el modelo I ( $29.2 \mu\text{g}/\text{m}^3$ ). En ambos modelos se observa una disminución en los coeficientes de determinación ( $r^2$ ) y un aumento (modelo I) y disminución (modelo II) de la RECM, al pasar de la etapa de entrenamiento a la de validación. Similares resultados se obtuvieron en Abdullah *et al.* (2020), donde generaron modelos de regresión lineal múltiple para el pronóstico de  $\text{PM}_{10}$  en Malasia. La **Tabla 46** presenta una comparación con estudios similares de modelos de regresión lineal múltiple para el pronóstico de  $\text{PM}_{10}$ .

**Tabla 40: Modelos de regresión lineal múltiple para el pronóstico horario de  $\text{PM}_{10}$  e indicadores de desempeño**

	Modelo	r	$r^2$	$r^2_{\text{ajustado}}$	ECM	RECM
I	E $\text{PM}_{10}=19.267-0.193\text{PM}_{10\text{CDM}}+0.776\text{PM}_{10\text{STA}}-$ $0.036\text{PM}_{10\text{VMT}}+0.420\text{PM}_{10\text{HCH}}+0.169\text{PM}_{10\text{SMP}}$	0.77	0.60	0.60	829.3	28.8
	V $\text{PM}_{10\text{pronosticado}}=47.154+0.586\text{PM}_{10\text{observado}}$	0.77	0.59	0.59	854.0	29.2
II	E $\text{PM}_{10}=231.519+0.813\text{PM}_{2,5}+0.021\text{CO}+$ $0.476\text{NO}_2-1.405\text{O}_3-3.333\text{T}-1.294\text{HR}-$ $5.003\text{WS}-0.053\text{WD}+0.047\text{RS}$	0.70	0.49	0.49	795.3	28.2
	V $\text{PM}_{10\text{pronosticado}}=58.264+0.496\text{PM}_{10\text{observado}}$	0.70	0.48	0.48	764.8	27.7

\*E: entrenamiento; V: validación

#### 4.3.2 Pronóstico del material particulado $\text{PM}_{2,5}$

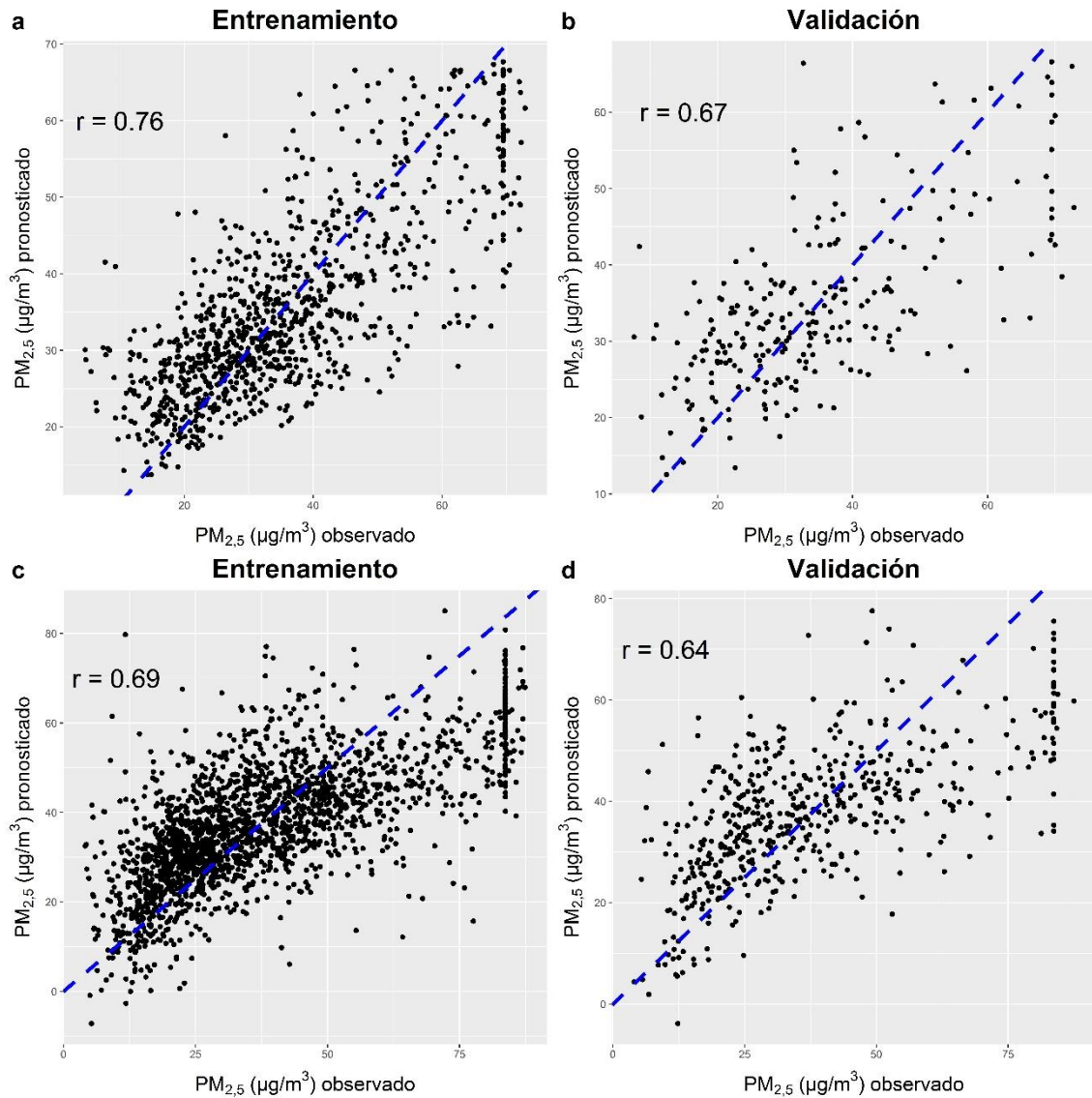
En la **Figura 31** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de  $\text{PM}_{2,5}$ , en la etapa de entrenamiento y validación. En la **Figura 31 a y b** se presenta el modelo de predicción para el  $\text{PM}_{2,5}$ , implementado a partir de las concentraciones de  $\text{PM}_{2,5}$  del resto de estaciones de calidad de aire. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.76, el cual es mayor al de la etapa de validación que tiene un valor de 0.67. Así también, en la **Figura 31 c y d** se presenta el modelo de predicción para el  $\text{PM}_{2,5}$ , implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. El coeficiente de correlación de Pearson entre los

valores observados y pronosticados para la etapa de entrenamiento es 0.69, el cual es mayor al de la etapa de validación el cual tiene un valor de 0.64.

El mejor modelo para el pronóstico de  $PM_{2.5}$  fue el modelo I. A pesar de tener valores de coeficientes de Pearson moderados, los modelos generados a partir de las variables meteorológicas no aportaron mayor captura de información, resultados similares se encontraron en (Zhang y Ding 2017). La **Tabla 41** presenta los indicadores de desempeño de los modelos generados.

Para el primer modelo de pronóstico de  $PM_{2.5}$  los predictores significativos fueron las concentraciones de  $PM_{2.5}$  de las estaciones STA ( $PM_{2.5STA}$ ), HCH( $PM_{2.5HCH}$ ), SJL( $PM_{2.5SJL}$ ) y PPD( $PM_{2.5PPD}$ ). Las concentraciones pronosticadas de  $PM_{2.5}$  para la estación ATE aumentan en 0.404, 0.299, 0.200 y 0.104 unidades cuando las variables  $PM_{2.5STA}$ ,  $PM_{2.5HCH}$ ,  $PM_{2.5SJL}$  y  $PM_{2.5PPD}$  aumentan en una unidad, respectivamente. Mientras tanto, el segundo modelo fue influenciado por ocho variables diferentes, es decir,  $PM_{10}$ , CO,  $NO_2$ ,  $SO_2$ ,  $O_3$ , TEMP, HR, y WS. Donde se observó que existe un aumento de 0.199, 0.003, 0.305, 0.241, 0.371, 0.507, 0.786 y una disminución de 4.401 unidades de  $PM_{2.5}$  pronosticado por unidad en aumento de  $PM_{10}$ , CO,  $NO_2$ ,  $SO_2$ ,  $O_3$ , T y WS, respectivamente (**Tabla 41**).

En la etapa de entrenamiento de los modelos, los coeficientes de determinación ( $r^2$ ) variaron de 0.58 a 0.48 para los modelos I y II, respectivamente. El menor valor de la raíz del error cuadrático medio (RECM), se presentó en el modelo I ( $10.0 \mu\text{g}/\text{m}^3$ ) respecto del modelo II ( $14.2 \mu\text{g}/\text{m}^3$ ). Mientras que, en la etapa de validación los coeficientes de determinación ( $r^2$ ) variaron de 0.45 a 0.41 para los modelos I y II, respectivamente; asimismo, el menor valor de la RECM se obtuvo en el modelo I ( $11.4 \mu\text{g}/\text{m}^3$ ) en comparación con el modelo II ( $15.4 \mu\text{g}/\text{m}^3$ ). Se observa que, en los modelos I y II hubo una disminución en los coeficientes de determinación ( $r^2$ ) y un aumento de la RECM al pasar de la etapa de entrenamiento al de validación. Estudios anteriores obtuvieron coeficientes de determinación ( $r^2$ ) en el rango de 0.40 a 0.72 y la RECM en el rango de 0.27 a  $29.6 \mu\text{g}/\text{m}^3$  para el pronóstico de  $PM_{2.5}$  a través de modelos de regresión lineal múltiple en Hong Kong, Malasia y México (Ordieres *et al.* 2005, Elbayoumi *et al.* 2015, Zhang y Ding 2017) (ver **Tabla 46**).



**Figura 31: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de PM<sub>2,5</sub> para la estación ATE**

**Tabla 41: Modelos de regresión lineal múltiple para el pronóstico horario de PM<sub>2,5</sub> e indicadores de desempeño**

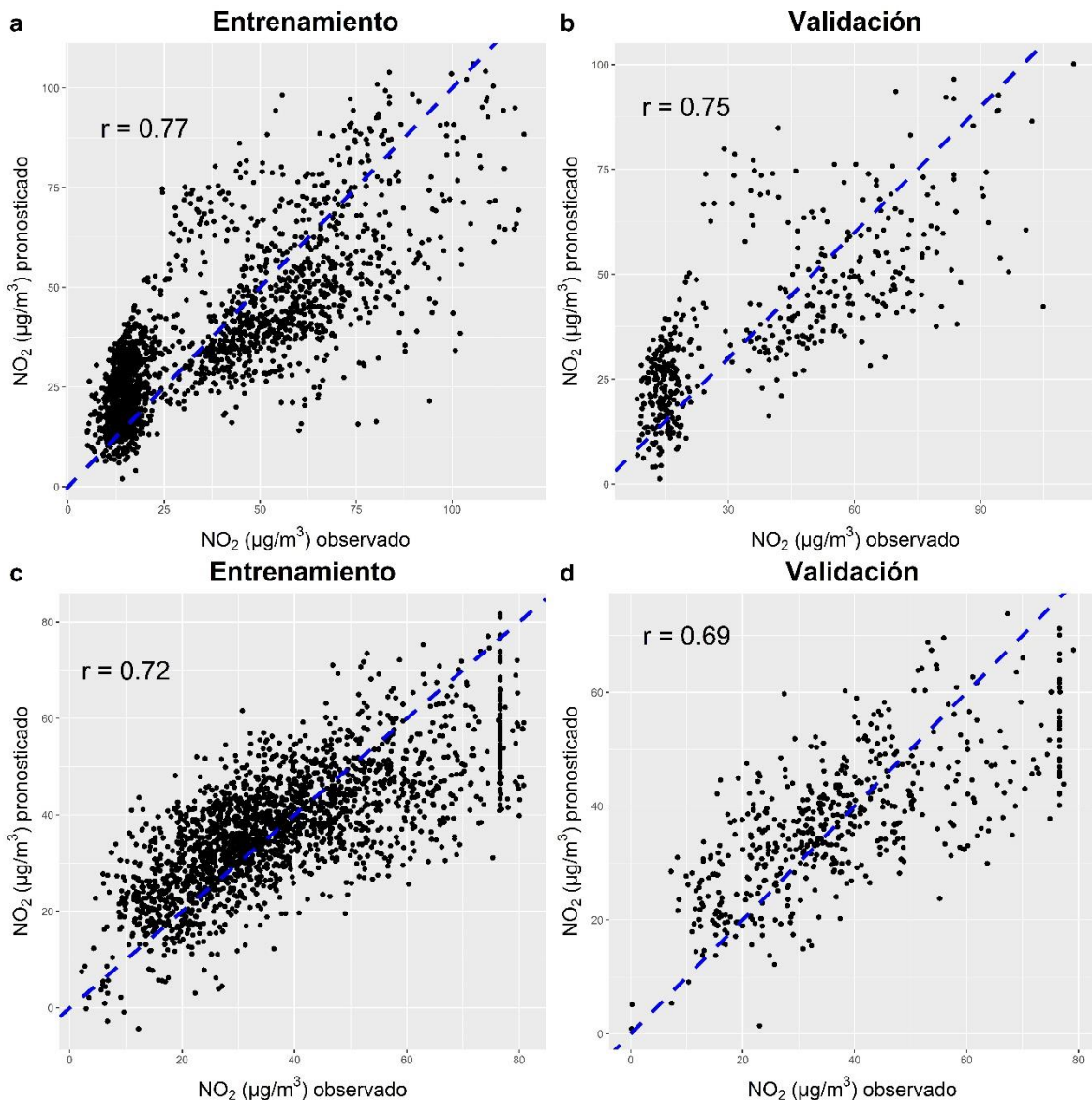
	Modelo	r	r <sup>2</sup>	r <sup>2</sup> <sub>ajustado</sub>	ECM	RECM
I	$PM_{2,5} = 4.410 + 0.404PM_{2,5STA} + 0.299PM_{2,5HCH} + 0.200PM_{2,5SJL} + 0.104PM_{2,5PPD}$	0.8	0.6	0.58	100.3	10.0
	V	$PM_{2,5\text{pronosticado}} = 17.937 + 0.484PM_{2,5\text{observado}}$	0.7	0.5	0.45	129.0
II	$PM_{2,5} = -78.580 + 0.199PM_{10} + 0.03CO + 0.305NO_2 + 0.241SO_2 + 0.371O_3 + 0.507T + 0.786HR - 4.401WS$	0.7	0.5	0.48	201.4	14.2
	V	$PM_{2,5\text{pronosticado}} = 21.861 + 0.426PM_{2,5\text{observado}}$	0.6	0.4	0.41	235.7

\*E: entrenamiento; V: validación

### 4.3.3 Pronóstico del dióxido de nitrógeno (NO<sub>2</sub>)

En la **Figura 32** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de NO<sub>2</sub>, en la etapa de entrenamiento y validación. En la **Figura 32 a y b** se presenta el modelo de predicción para el NO<sub>2</sub>, implementado a partir de las concentraciones de NO<sub>2</sub> del resto de estaciones de calidad de aire. Los coeficientes de correlación de Pearson entre los valores observados y pronosticados, para las etapas de entrenamiento y validación son de 0.77 y 0.75, respectivamente. Así también, en la **Figura 32 c y d** se presenta el modelo de predicción para el NO<sub>2</sub>, implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.72, el cual es mayor al de la etapa de validación, que tiene un valor de 0.69. El mejor modelo para el pronóstico de NO<sub>2</sub> fue el modelo I. A pesar de tener valores de coeficientes de Pearson moderados, los modelos generados a partir de las variables meteorológicas no aportaron mayor captura de información, resultados similares se encontraron en (Rahimi *et al.* 2017). En la **Tabla 42** se presentan los indicadores de desempeño de los modelos generados para el NO<sub>2</sub>.

Para el primer modelo de pronóstico de NO<sub>2</sub> los predictores significativos fueron las concentraciones de NO<sub>2</sub> de las estaciones SBJ(NO<sub>2</sub><sub>SBJ</sub>), CDM(NO<sub>2</sub><sub>CDM</sub>), STA(NO<sub>2</sub><sub>STA</sub>), VMT(NO<sub>2</sub><sub>VMT</sub>), HCH(NO<sub>2</sub><sub>HCH</sub>), SJL(NO<sub>2</sub><sub>SJL</sub>) y PPD(NO<sub>2</sub><sub>PPD</sub>). Las concentraciones pronosticadas de NO<sub>2</sub> para la estación ATE aumentan en 0.256, 1.506, 0.209 y 0.425 unidades cuando las variables NO<sub>2</sub><sub>SBJ</sub>, NO<sub>2</sub><sub>VMT</sub>, NO<sub>2</sub><sub>HCH</sub> y NO<sub>2</sub><sub>SJL</sub> aumentan en una unidad, respectivamente; y disminuyen en 0.576, 0.318 y 0.094 unidades cuando las variables NO<sub>2</sub><sub>CDM</sub>, NO<sub>2</sub><sub>STA</sub> y NO<sub>2</sub><sub>PPD</sub>, aumentan en una unidad individualmente. Por otro lado, el segundo modelo fue influenciado por nueve variables diferentes, es decir, PM<sub>10</sub>, PM<sub>2,5</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, T, HR, WS y RS. Se observó que, existe un aumento de 0.091, 0.221, 0.012, 2.013, 2.057, 0.018 y una disminución de 0.350, 0.463, 0.261 unidades de NO<sub>2</sub> pronosticado por unidad en aumento de PM<sub>10</sub>, PM<sub>2,5</sub>, CO, T, WS, RS, SO<sub>2</sub>, O<sub>3</sub> y HR, respectivamente (**Tabla 42**).



**Figura 32: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de NO<sub>2</sub> para la estación ATE**

En la etapa de entrenamiento de los modelos, los coeficientes de determinación ( $r^2$ ) variaron de 0.60 a 0.52 para los modelos I y II, respectivamente. El menor valor de la raíz del error cuadrático medio (RECM), se presentó en el modelo II (11.9 µg/m<sup>3</sup>) respecto del modelo I (16.3 µg/m<sup>3</sup>). Mientras que, en la etapa de validación los coeficientes de determinación ( $r^2$ ) variaron de 0.57 a 0.47 para los modelos I y II, respectivamente; asimismo, el menor valor de la RECM se obtuvo en el modelo II (12.8 µg/m<sup>3</sup>) en comparación con el modelo I (16.2 µg/m<sup>3</sup>). Se observa que, en los modelos I y II hubo una disminución en los coeficientes de determinación ( $r^2$ ), mientras la RECM aumentó para el modelo I y disminuyó para el modelo II, al pasar de la etapa de entrenamiento al de validación. En estudios anteriores se obtuvieron coeficientes de determinación ( $r^2$ ) de 0.41 a 0.67 y RECM de 3.6 a 35.1 µg/m<sup>3</sup> en el



pronóstico de NO<sub>2</sub> a través de modelos de regresión lineal múltiple en Irán y Hong Kong (ver **Tabla 46**) (Rahimi *et al.* 2017, Zhang y Ding 2017).

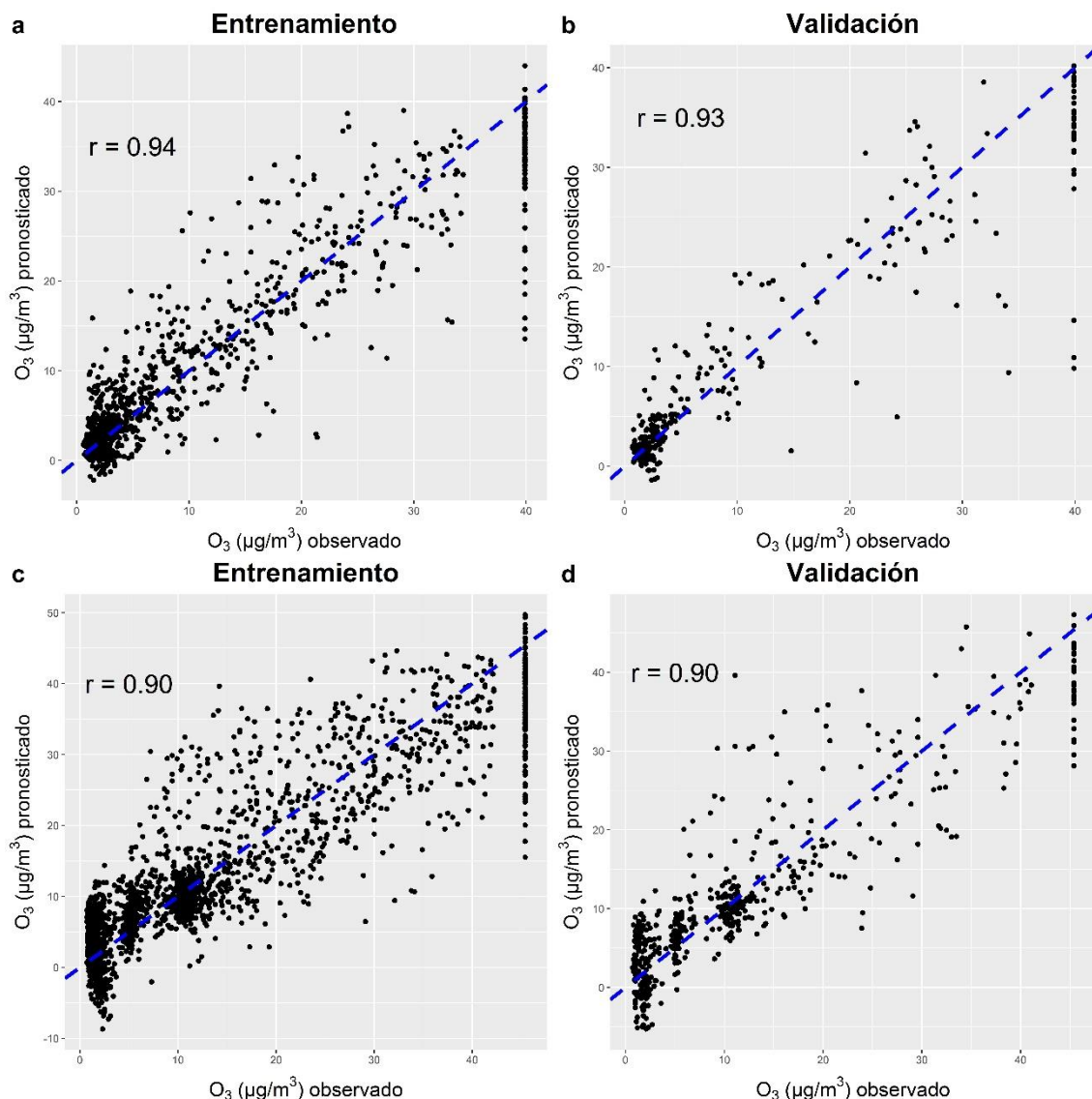
**Tabla 42: Modelos de regresión lineal múltiple para el pronóstico horario de NO<sub>2</sub> e indicadores de desempeño**

	Modelo	r	r <sup>2</sup>	r <sup>2</sup> <sub>ajustado</sub>	ECM	RECM
	NO <sub>2</sub> =9.529+0.256NO <sub>2</sub> SBJ-0.576NO <sub>2</sub> CDM-					
I	E 0.318NO <sub>2</sub> STA+1.506NO <sub>2</sub> VMT+0.209NO <sub>2</sub> HCH	0.8	0.60	0.60	266.2	16.3
	+0.425NO <sub>2</sub> SJL-0.094NO <sub>2</sub> PPD					
V	NO <sub>2</sub> pronosticado=16.020+0.592 NO <sub>2</sub> observado	0.8	0.57	0.57	262.9	16.2
	NO <sub>2</sub> = -6.437+0.091PM <sub>10</sub> +0.221PM <sub>2,5</sub> +					
II	E 0.012CO-0.350SO <sub>2</sub> -0.463O <sub>3</sub> +2.013T-	0.7	0.52	0.52	142.8	11.9
	0.261HR+2.057WS+0.018RS					
V	NO <sub>2</sub> pronosticado=19.596+0.470NO <sub>2</sub> observado	0.7	0.47	0.47	163.8	12.8

\*E: entrenamiento; V: validación

#### 4.3.4 Pronóstico del ozono troposférico (O<sub>3</sub>)

En la **Figura 33** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de O<sub>3</sub>, en la etapa de entrenamiento como en la de validación. En la **Figura 33 a y b** se presenta el modelo de predicción para el O<sub>3</sub>, implementado a partir de las concentraciones de O<sub>3</sub> del resto de estaciones de calidad de aire. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.94, el cual es ligeramente mayor al de la etapa de validación con un valor de 0.93. Así también, en la **Figura 33 c y d** se presenta el modelo de predicción para el O<sub>3</sub>, implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. Los coeficientes de correlación de Pearson entre los valores observados y pronosticados, para ambas etapas son similares con un valor de 0.90. El mejor modelo para el pronóstico de O<sub>3</sub> fue el modelo I. A pesar de tener valores de coeficientes de Pearson altos, los modelos generados a partir de las variables meteorológicas no aportaron mayor captura de información, resultados similares se encontraron en (Abdullah *et al.* 2019). En la **Tabla 43** se presentan los indicadores de desempeño de los modelos generados.



**Figura 33: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de O<sub>3</sub> para la estación ATE**

Para el primer modelo de pronóstico de O<sub>3</sub> los predictores significativos fueron las concentraciones de O<sub>3</sub> de las estaciones STA(O<sub>3</sub>STA), VMT(O<sub>3</sub>VMT), HCH(O<sub>3</sub>HCH), SJL(O<sub>3</sub>SJL), CRB(O<sub>3</sub>CRB) y PPD(O<sub>3</sub>PPD). Las concentraciones pronosticadas de O<sub>3</sub> para la estación ATE aumentan en 0.353, 0.219, 0.164 y 0.238 unidades cuando las variables O<sub>3</sub>STA, O<sub>3</sub>HCH, O<sub>3</sub>SJL y O<sub>3</sub>PPD aumentan en una unidad, respectivamente; y disminuyen en 0.041 y 0.115 unidades cuando las variables O<sub>3</sub>VMT y O<sub>3</sub>CRB, aumentan en una unidad individualmente. Mientras tanto, el segundo modelo fue influenciado por diez variables diferentes, es decir, PM<sub>10</sub>, PM<sub>2,5</sub>, CO, NO<sub>2</sub>, SO<sub>2</sub>, T, HR, WS, WD y RS. Se observó que, existe un aumento de 0.075, 0.007, 2.181, 0.033 y una disminución de 0.062, 0.118, 0.138, 0.925, 0.440, 0.016 unidades de O<sub>3</sub> pronosticado por unidad en aumento de PM<sub>2,5</sub>, CO, WS, RS, PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, T, HR y WD, respectivamente (**Tabla 43**).

En la etapa de entrenamiento de los modelos, los coeficientes de determinación ( $r^2$ ) variaron de 0.88 a 0.80 para los modelos I y II, respectivamente. El menor valor de la raíz del error cuadrático medio (RECM), se presentó en el modelo I ( $4.2 \mu\text{g}/\text{m}^3$ ) respecto del modelo II ( $6.0 \mu\text{g}/\text{m}^3$ ). Mientras que, en la etapa de validación los coeficientes de determinación ( $r^2$ ) variaron de 0.86 a 0.80 para los modelos I y II, respectivamente; asimismo, el menor valor de la RECM se obtuvo en el modelo I ( $5.1 \mu\text{g}/\text{m}^3$ ) en comparación con el modelo II ( $5.5 \mu\text{g}/\text{m}^3$ ). Se observa que, en el modelo I hubo una disminución en los coeficientes de determinación ( $r^2$ ); mientras la RECM aumentó para el modelo I y disminuyó para el modelo II, al pasar de la etapa de entrenamiento al de validación. En estudios recientes se obtuvieron coeficientes de determinación ( $r^2$ ) en el rango de 0.32 a 0.81 y RECM en el rango de 14.5 a  $31.8 \mu\text{g}/\text{m}^3$  en el pronóstico de  $\text{O}_3$  a través de modelos de regresión lineal múltiple en Malasia y Hong Kong (ver **Tabla 46**) (Abdullah *et al.* 2019, Napi *et al.* 2020, Zhang y Ding 2017).

**Tabla 43: Modelos de regresión lineal múltiple para el pronóstico horario de  $\text{O}_3$  e indicadores de desempeño**

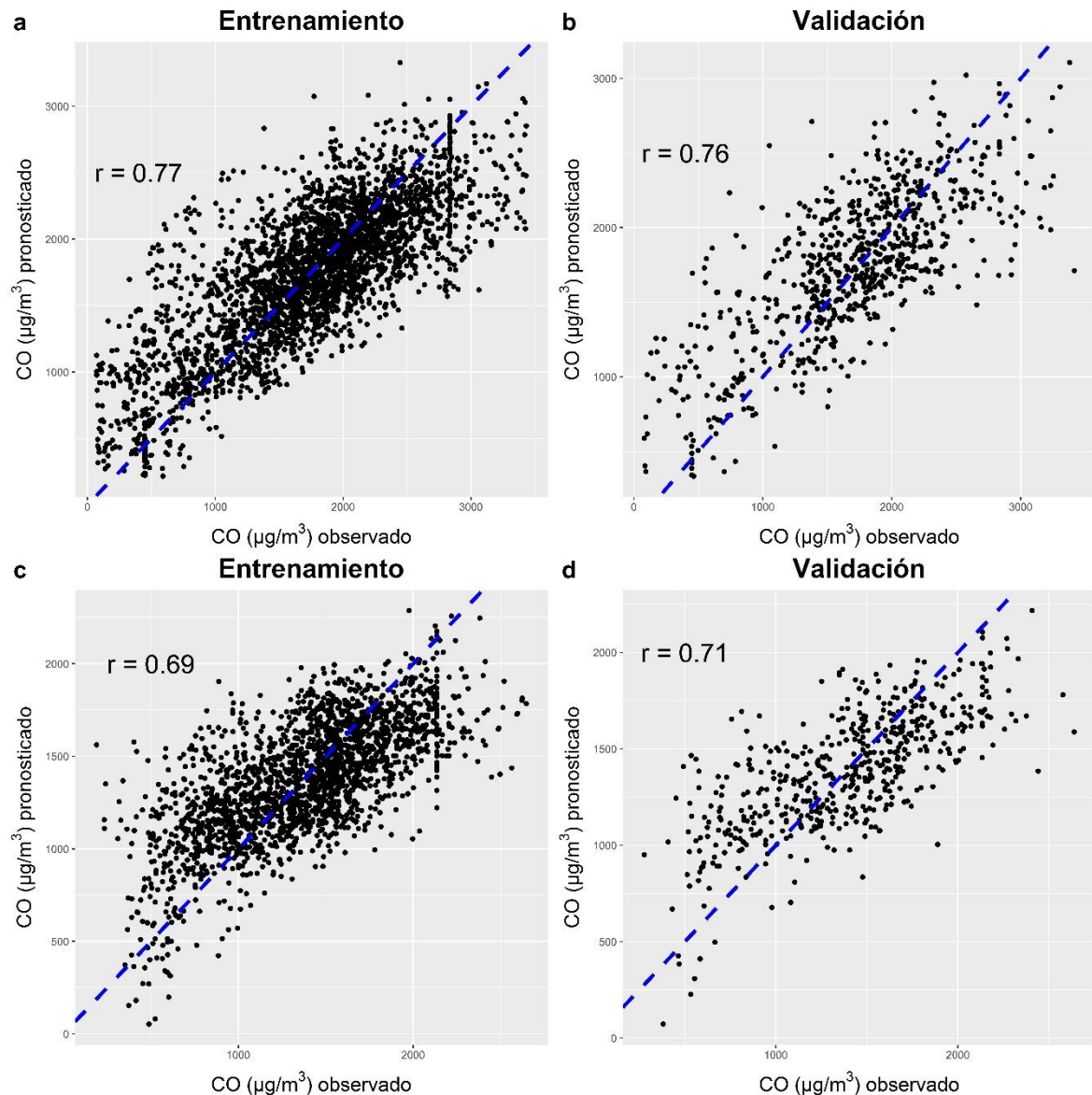
	Modelo	r	$r^2$	$r^2_{\text{ajustado}}$	ECM	RECM
	$\text{O}_3 = -0.076 + 0.353\text{O}_{3\text{STA}}$					
I	E $0.041\text{O}_{3\text{VMT}} + 0.219\text{O}_{3\text{HCH}} + 0.164\text{O}_{3\text{SIL}} -$ $0.115\text{O}_{3\text{CRB}} + 0.238\text{O}_{3\text{PPD}}$	0.94	0.88	0.88	17.5	4.2
	V $\text{O}_{3\text{pronosticado}} = 1.334 + 0.817\text{O}_{3\text{observado}}$	0.93	0.86	0.86	25.5	5.1
	$\text{O}_3 = 63.459 - 0.062\text{PM}_{10} + 0.075\text{PM}_{2.5} +$					
II	E $0.007\text{CO} - 0.118\text{NO}_2 - 0.138\text{SO}_2 - 0.925\text{T} -$ $0.440\text{HR} + 2.181\text{WS} - 0.016\text{WD} + 0.033\text{RS}$	0.90	0.80	0.80	36.1	6.0
	V $\text{O}_{3\text{pronosticado}} = 2.198 + 0.827\text{O}_{3\text{observado}}$	0.90	0.80	0.80	30.8	5.5

\*E: entrenamiento; V: validación

#### 4.3.5 Pronóstico del monóxido de carbono (CO)

En la **Figura 34** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de CO, en la etapa de entrenamiento y validación. En la **Figura 34 a y b** se presenta el modelo de predicción para el CO, implementado a partir de las concentraciones de CO del resto de estaciones de calidad de aire. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.77, el cual es ligeramente mayor al de la etapa de validación, que tiene

un valor de 0.76. Así también, en la **Figura 34 c y d** se presenta el modelo de predicción para el CO, implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.69, el cual es menor al de la etapa de validación que presenta un valor de 0.71. El mejor modelo para el pronóstico de CO fue el modelo I. En la **Tabla 44** se presentan los indicadores de desempeño de los modelos generados para el pronóstico de CO.



**Figura 34: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de CO para la estación ATE**

Para el primer modelo de pronóstico de CO los predictores significativos fueron las concentraciones de CO de las estaciones SBJ( $\text{CO}_{\text{SBJ}}$ ), CDM( $\text{CO}_{\text{CDM}}$ ), STA( $\text{CO}_{\text{STA}}$ ), VMT( $\text{CO}_{\text{VMT}}$ ), HCH( $\text{CO}_{\text{HCH}}$ ), SJL( $\text{CO}_{\text{SJL}}$ ), SMP( $\text{CO}_{\text{SMP}}$ ), CRB( $\text{CO}_{\text{CRB}}$ ) y PPD( $\text{CO}_{\text{PPD}}$ ). Las

concentraciones pronosticadas de CO para la estación ATE aumentan en 0.103, 0.864, 0.503, 0.537 y 0.469 unidades cuando las variables  $CO_{SBJ}$ ,  $CO_{STA}$ ,  $CO_{VMT}$ ,  $CO_{HCH}$  y  $CO_{PPD}$  aumentan en una unidad, respectivamente; y disminuyen en 0.482, 0.361, 0.236 y 0.203 unidades cuando las variables  $CO_{CDM}$ ,  $CO_{SIL}$ ,  $CO_{SMP}$  y  $CO_{CRB}$ , aumentan en una unidad individualmente. Mientras tanto, el segundo modelo fue influenciado por ocho variables diferentes, es decir,  $PM_{10}$ ,  $PM_{2.5}$ ,  $NO_2$ ,  $SO_2$ ,  $O_3$ , T, HR y RS. Existe un aumento de 2.764, 1.287, 8.778, 5.423, 20.118, 14.714 y una disminución de 21.498, 0.653 unidades de CO pronosticado por unidad en aumento de  $PM_{10}$ ,  $PM_{2.5}$ ,  $NO_2$ ,  $SO_2$ ,  $O_3$ , HR, T y RS, respectivamente (**Tabla 44**).

**Tabla 44: Modelos de regresión lineal múltiple para el pronóstico horario de CO e indicadores de desempeño**

	Modelo	r	r <sup>2</sup>	r <sup>2</sup> <sub>ajustado</sub>	ECM	RECM
I	CO=197.372+0.103CO <sub>SBJ</sub> -0.482CO <sub>CDM</sub> + 0.864CO <sub>STA</sub> +0.503CO <sub>VMT</sub> +0.537CO <sub>HCH</sub> - 0.361CO <sub>SIL</sub> -0.236CO <sub>SMP</sub> -0.203CO <sub>CRB</sub> + 0.469CO <sub>PPD</sub>	0.77	0.60	0.60	185454.5	430.6
	V CO <sub>pronosticado</sub> =744.534+0.587CO <sub>observado</sub>	0.76	0.57	0.57	187483.9	433.0
	CO=-334.094+2.764PM <sub>10</sub> +1.287PM <sub>2.5</sub> + 8.778NO <sub>2</sub> +5.423SO <sub>2</sub> +20.118O <sub>3</sub> - 21.498T+14.714HR-0.653RS	0.69	0.47	0.47	107108.0	327.3
II	V CO <sub>pronosticado</sub> =730.537+0.474CO <sub>observado</sub>	0.71	0.51	0.51	103055.5	321.0

\*E: entrenamiento; V: validación

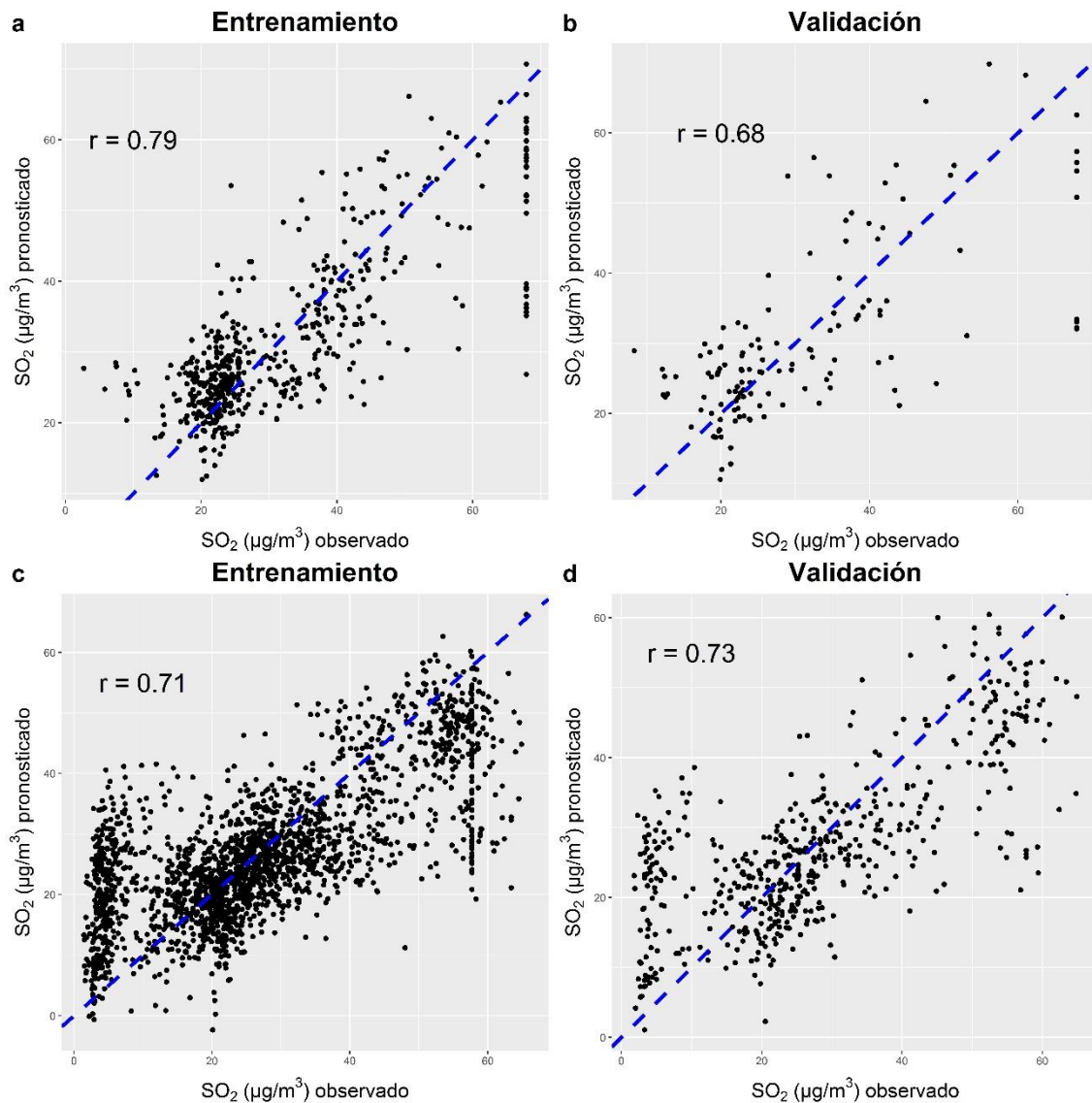
En la etapa de entrenamiento de los modelos, los coeficientes de determinación ( $r^2$ ) variaron de 0.60 a 0.47 para los modelos I y II, respectivamente. El menor valor de la raíz del error cuadrático medio (RECM), se presentó en el modelo II (327.3  $\mu\text{g}/\text{m}^3$ ) con respecto al modelo I (430.6  $\mu\text{g}/\text{m}^3$ ). Mientras que, en la etapa de validación los coeficientes de determinación ( $r^2$ ) variaron de 0.57 a 0.51 para los modelos I y II, respectivamente; asimismo, el menor valor de la RECM se obtuvo en el modelo II (321.0  $\mu\text{g}/\text{m}^3$ ) en comparación con el modelo I (433  $\mu\text{g}/\text{m}^3$ ). En el modelo I se observa una disminución en los coeficientes de determinación ( $r^2$ ) y un aumento en la RECM; mientras que, en el modelo II, se observa un aumento en el  $r^2$  y una disminución en la RECM, al pasar de la etapa de entrenamiento a la de validación. En un estudio reciente se obtuvo un coeficiente de determinación ( $r^2$ ) de alrededor de 0.01 y

una RECM de  $11.747 \mu\text{g}/\text{m}^3$  en el pronóstico de CO a través de un modelo de regresión lineal múltiple en Irán (Shams *et al.* 2020).

#### 4.3.6 Pronóstico del dióxido de azufre (SO<sub>2</sub>)

En la **Figura 35** se presentan los diagramas de dispersión de los dos modelos implementados para el pronóstico horario de SO<sub>2</sub>, en la etapa de entrenamiento como en la de validación. En la **Figura 35 a y b** se presenta el modelo de predicción para el SO<sub>2</sub>, implementado a partir de las concentraciones de SO<sub>2</sub> del resto de estaciones de calidad de aire. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.80, el cual es mayor al de la etapa de validación con un valor de 0.69. Así también, en la **Figura 35 c y d** se presenta el modelo de predicción para el SO<sub>2</sub>, implementado a partir de la concentración de los contaminantes y los valores de las variables meteorológicas de la misma estación. El coeficiente de correlación de Pearson entre los valores observados y pronosticados para la etapa de entrenamiento es 0.71, el cual es menor al de la etapa de validación, que tiene un valor de 0.73. El mejor modelo para el pronóstico de SO<sub>2</sub> fue el modelo II. El modelo generado a partir de las variables meteorológicas aportó mayor captura de información, de acuerdo con los valores de coeficientes de Pearson moderados en la etapa de validación, resultados similares se encontraron en (Zhang y Ding 2017). En la **Tabla 45** se presentan los indicadores de desempeño de los modelos generados.

Para el primer modelo de pronóstico de SO<sub>2</sub> los predictores significativos fueron las concentraciones de SO<sub>2</sub> de las estaciones SBJ(SO<sub>2</sub><sub>SBJ</sub>), CDM(SO<sub>2</sub><sub>CDM</sub>), STA(SO<sub>2</sub><sub>STA</sub>), SJL(SO<sub>2</sub><sub>SJL</sub>), SMP(SO<sub>2</sub><sub>SMP</sub>), CRB(SO<sub>2</sub><sub>CRB</sub>) y PPD(SO<sub>2</sub><sub>PPD</sub>). Las concentraciones pronosticadas de SO<sub>2</sub> para la estación ATE aumentan en 1.144, 0.751, 0.889, 1.997 y 0.212 unidades cuando las variables SO<sub>2</sub><sub>SBJ</sub>, SO<sub>2</sub><sub>STA</sub>, SO<sub>2</sub><sub>SJL</sub>, SO<sub>2</sub><sub>SMP</sub> y SO<sub>2</sub><sub>PPD</sub> aumentan en una unidad, respectivamente; y disminuyen en 2.977 y 1.330 unidades cuando las variables SO<sub>2</sub><sub>CDM</sub> y SO<sub>2</sub><sub>CRB</sub>, aumentan en una unidad individualmente. Mientras tanto, el segundo modelo fue influenciado por nueve variables diferentes, es decir, PM<sub>2,5</sub>, CO, NO<sub>2</sub>, O<sub>3</sub>, T, HR, WS, WD y RS. Existe un aumento de 0.165, 0.007, 2.559, 0.037, 0.0153 y una disminución de 0.299, 0.501, 0.298, 5.813 unidades de SO<sub>2</sub> pronosticado por unidad en aumento de PM<sub>2,5</sub>, CO, T, WD, RS, NO<sub>2</sub>, O<sub>3</sub>, HR y WS, respectivamente (**Tabla 45**).



**Figura 35: Diagramas de dispersión de los modelos de pronóstico de concentraciones horarias de SO<sub>2</sub> para la estación ATE**

Para el modelo I los coeficientes de determinación ( $r^2$ ) varían de 0.63 a 0.46, para las etapas de entrenamiento y validación, mientras que, el  $r^2$  ajustado presentaron los mismos valores para ambas etapas del modelamiento. Los errores cuadráticos medios (ECM), oscilan de 73.3 a 118.0 y la raíz del error cuadrático medio (RECM) varía de 8.6 a 10.9  $\mu\text{g}/\text{m}^3$ , para ambas etapas del pronóstico. Asimismo, para el modelo II los coeficientes de determinación ( $r^2$ ) varían de 0.51 a 0.53, para las etapas de entrenamiento y validación, respectivamente; mientras que, el  $r^2$  ajustado presentó los mismos valores para ambas etapas del pronóstico. Los errores cuadráticos medios (ECM), varían de 125.6 a 127.4 y la raíz del error cuadrático medio (RECM), oscila de 11.2 a 11.3  $\mu\text{g}/\text{m}^3$ , para ambas etapas del pronóstico. Similares resultados se obtuvieron en Zhang y Ding (2017) donde implementaron modelos de regresión lineal múltiple para el pronóstico de SO<sub>2</sub> en Hong Kong (ver **Tabla 46**).

**Tabla 45: Modelos de regresión lineal múltiple para el pronóstico horario de SO<sub>2</sub> e indicadores de desempeño**

	Modelo	r	r <sup>2</sup>	r <sup>2</sup> <sub>ajustado</sub>	ECM	RECM
	$SO_2=9.599+1.144SO_{2SBJ}-$					
I	E $2.977SO_{2CDM}+0.751SO_{2STA}+0.889SO_{2S JL}+$	0.79	0.63	0.63	73.3	8.6
	$1.997SO_{2SMP}-1.330SO_{2CRB}+0.212SO_{2PPD}$					
V	$SO_{2pronosticado}=12.603+0.587SO_{2observado}$	0.68	0.46	0.46	118.0	10.9
	$SO_2=-1.832+0.165PM_{2,5}+0.007CO-$					
II	E $0.299NO_2-0.501O_3+2.559T-0.298HR-$	0.71	0.51	0.51	125.6	11.2
	$5.813WS+0.037WD+0.0153RS$					
V	$SO_{2pronosticado}=13.091+0.518SO_{2observado}$	0.73	0.53	0.53	127.4	11.3

\*E: entrenamiento; V: validación

**Tabla 46: Comparación con estudios similares de modelos de regresión lineal múltiple**

Fuente	País	Contaminante	R <sup>2</sup>	RECM
Abdullah <i>et al.</i> (2020)	Malasia	PM <sub>10</sub>	0.13-0.64	126.73-164.98
Abdullah <i>et al.</i> (2019)	Malasia	PM <sub>10</sub>	0.55-0.71	10.6-28
Ceylan y Bulkan (2018)	Turquía	PM <sub>10</sub>	0.32	N.R.
Abdullah <i>et al.</i> (2018)	Malasia	PM <sub>10</sub>	0.53-0.65	33.278
Ul-Saufie <i>et al.</i> (2013)	Malasia	PM <sub>10</sub>	0.35-0.62	14.23-23.03
Zhang y Ding (2017)	Hong Kong	PM <sub>2,5</sub>	0.5-0.7	16.4-29.6
Elbayoumi <i>et al.</i> (2015)	Malasia	PM <sub>2,5</sub>	0.54-0.72	1.3-1.7
Ordieres <i>et al.</i> (2005)	México	PM <sub>2,5</sub>	0.40	0.27
Rahimi <i>et al.</i> (2017)	Irán	NO <sub>2</sub>	0.41	3.6
Zhang y Ding (2017)	Hong Kong	NO <sub>2</sub>	0.52-0.67	19-35.1
Napi <i>et al.</i> (2020)	Malasia	O <sub>3</sub>	0.32	N.R.
Abdullah <i>et al.</i> (2019)	Malasia	O <sub>3</sub>	0.79-0.81	N.R.
Zhang y Ding (2017)	Hong Kong	O <sub>3</sub>	0.54-0.64	14.5-31.8
Shams <i>et al.</i> (2020)	Irán	CO	0.01	11.747
Zhang y Ding (2017)	Hong Kong	SO <sub>2</sub>	0.54-0.74	7.9-18.1

N.R.: no registra



Finalmente, se presentan en las **Tablas 47 y 48**, los modelos de regresión lineal múltiple para el pronóstico horario de PM<sub>10</sub> y PM<sub>2,5</sub>, respectivamente, en las estaciones VMT (zona sur), CDM (zona centro) y PPD (zona norte) del AMLC (**Anexo 3**).

**Tabla 47: Modelos de regresión lineal múltiple para el pronóstico horario de PM<sub>10</sub> e indicadores de desempeño para las estaciones VMT, CDM y PPD**

Estación	Modelo	r	r <sup>2</sup>	r <sup>2</sup> ajustado	ECM	RECM		
<b>VMT</b>	I E	PM <sub>10</sub> =-25.514-0.143PM <sub>10ATE</sub> + 0.559PM <sub>10SBJ</sub> -0.247PM <sub>10CDM</sub> + 0.536PM <sub>10STA</sub> +0.196PM <sub>10HCH</sub> - 0.494PM <sub>10SIL</sub> +0.369PM <sub>10SMP</sub> + 0.699PM <sub>10CRB</sub> +0.496PM <sub>10PPD</sub>	0.63	0.40	0.40	3596.8	60.0	
		V	PM <sub>10pronosticado</sub> =74.894+ 0.374PM <sub>10observado</sub>	0.61	0.37	0.37	3777.8	61.5
			II E	PM <sub>10</sub> =247.059+1.850PM <sub>2,5</sub> + 0.079CO+1.860NO <sub>2</sub> +1.260SO <sub>2</sub> - 1.082O <sub>3</sub> +0.888T-2.799HR- 0.095WD-0.025RS	0.77	0.60	0.60	2448.5
	V	PM <sub>10pronosticado</sub> =55.256+ 0.593PM <sub>10observado</sub>	0.77	0.60	0.59	2476.8	49.8	
	<b>CDM</b>	I E	PM <sub>10</sub> =2.434-0.022PM <sub>10ATE</sub> + 0.317PM <sub>10SBJ</sub> +0.088PM <sub>10STA</sub> - 0.006PM <sub>10VMT</sub> +0.074PM <sub>10SIL</sub> + 0.166PM <sub>10SMP</sub> -0.063PM <sub>10CRB</sub> + 0.032PM <sub>10PPD</sub>	0.8	0.58	0.58	83.7	9.1
			V	PM <sub>10pronosticado</sub> =14.845+ 0.547PM <sub>10observado</sub>	0.7	0.55	0.55	94.0
II E		PM <sub>10</sub> =67.027+1.226PM <sub>2,5</sub> + 0.017CO+0.439NO <sub>2</sub> +0.706SO <sub>2</sub> - 1.341T-0.478HR-0.802WS+ 0.025WD	0.8	0.58	0.58	100.7	10.0	
V		PM <sub>10pronosticado</sub> =16.561+ 0.552PM <sub>10observado</sub>	0.7	0.56	0.56	108.2	10.4	
<b>PPD</b>		I E	PM <sub>10</sub> =29.264+0.525PM <sub>10CDM</sub> + 0.174PM <sub>10VMT</sub> +0.401PM <sub>10SIL</sub> - 0.090PM <sub>10SMP</sub> +0.070PM <sub>10CRB</sub>	0.6	0.37	0.37	1246.9	35.3
	V		PM <sub>10pronosticado</sub> =66.499+ 0.369PM <sub>10observado</sub>	0.6	0.35	0.35	1234.3	35.1
	II E	PM <sub>10</sub> =275.5+1.675PM <sub>2,5</sub> + 0.013CO+0.239NO <sub>2</sub> -0.146SO <sub>2</sub> - 0.568O <sub>3</sub> -0.750T-2.454HR+ 3.567WS-0.139WD-0.007RS	0.7	0.56	0.56	885.8	29.8	
	V	PM <sub>10pronosticado</sub> =47.655+ 0.554PM <sub>10observado</sub>	0.7	0.56	0.56	894.1	29.9	

\*E: entrenamiento; V: validación

**Tabla 48: Modelos de regresión lineal múltiple para el pronóstico horario de PM<sub>2,5</sub> e indicadores de desempeño para las estaciones VMT, CDM y PPD**

Estación	Modelo	r	r <sup>2</sup>	r <sup>2</sup> ajustado	ECM	RECM			
<b>VMT</b>	I	E	$PM_{2,5}=4.432+0.123PM_{2,5SBJ}+$ $0.133PM_{2,5CDM}+0.123PM_{2,5HCH}+$ $0.129PM_{2,5SJL}+0.167PM_{2,5SMP}+$ $0.233PM_{2,5PPD}$	0.63	0.39	0.39	94.3	9.7	
		V	$PM_{2,5pronosticado}=15.945+$ $0.404PM_{2,5observado}$	0.66	0.44	0.44	82.3	9.1	
	II	E	$PM_{2,5}=-14.310+0.065PM_{10}+$ $0.007CO+0.144NO_2+0.752SO_2+$ $0.205T+0.174HR-5.332WS+$ $0.024WD+0.018RS$	0.70	0.49	0.49	88.4	9.4	
		V	$PM_{2,5pronosticado}=14.001+$ $0.455PM_{2,5observado}$	0.67	0.45	0.45	97.0	9.9	
	<b>CDM</b>	I	E	$PM_{2,5}=2.912+0.216PM_{2,5SBJ}+$ $0.049PM_{2,5VMT}+0.062PM_{2,5HCH}+$ $0.372PM_{2,5SMP}$	0.67	0.45	0.45	33.1	5.8
			V	$PM_{2,5pronosticado}=8.759+$ $0.452PM_{2,5observado}$	0.67	0.44	0.44	30.9	5.6
II		E	$PM_{2,5}=-2.943+0.205PM_{10}+$ $0.004CO-0.098NO_2-0.135SO_2+$ $0.043O_3+0.106HR$	0.61	0.38	0.38	16.6	4.1	
		V	$PM_{2,5pronosticado}=7.451+$ $0.388PM_{2,5observado}$	0.62	0.38	0.38	16.1	4.0	
<b>PPD</b>	I	E	$PM_{2,5}=10.796+0.107 PM_{2,5ATE}+$ $0.191 PM_{2,5VMT}+0.088 PM_{2,5HCH}+$ $0.089 PM_{2,5SJL}+0.167 PM_{2,5SMP}+$ $0.245 PM_{2,5CRB}$	0.71	0.50	0.50	74.5	8.6	
		V	$PM_{2,5pronosticado}=19.988+$ $0.442PM_{2,5observado}$	0.62	0.39	0.38	87.8	9.4	
	II	E	$PM_{2,5}=-46.350+0.159PM_{10}+$ $0.001CO+0.086NO_2+0.122SO_2+$ $0.092O_3+0.245T+0.632HR-$ $4.736WS+0.018WD+0.023RS$	0.68	0.46	0.46	84.2	9.2	
		V	$PM_{2,5pronosticado}=17.379+$ $0.441PM_{2,5observado}$	0.67	0.45	0.45	88.5	9.4	

\*E: entrenamiento; V: validación

## V. CONCLUSIONES

1. El ACP y el ACJ se complementaron en la identificación de patrones espaciales de los contaminantes atmosféricos en el AMLC. Se identificaron 6 patrones de contaminación para el  $PM_{10}$  y  $PM_{2,5}$ , 5 para el  $NO_2$  y CO, 3 para el  $O_3$  y 7 para el  $SO_2$ . Los analizadores de contaminantes atmosféricos correspondientes a las regiones homogéneas se pueden instalar en zonas no monitoreadas, permitiendo así la ampliación de la REMCA. Las estaciones de la zona este presentan mayores concentraciones de  $PM_{10}$ ,  $PM_{2,5}$ ,  $NO_2$ , CO y  $SO_2$ ; mientras que, las estaciones en la zona centro, presentan menores concentraciones para los mismos contaminantes. El hecho que, los contaminantes presenten distribuciones espaciales similares es un buen indicador de que los niveles de concentración se determinan principalmente por las características topográficas y meteorológicas de cada zona en el AMLC.
2. Se determinaron los patrones espaciales de las variables meteorológicas en el AMLC, mediante el ACP y el ACJ. Se identificaron 2 patrones diferentes para la T, la HR y la RS; 5 patrones para la VV; y 7 patrones para la DV. Los patrones meteorológicos presentan zonas homogéneas para las variables T, HR y RS. En cambio, la DV y VV presentan mayor heterogeneidad. Las estaciones SMP, SBJ y CDM, presentan patrones espaciales similares de HR, las cuales se diferencian del resto de las estaciones de monitoreo. Para la T, la estación SMP se diferencia del resto de estaciones. Respecto a la RS, la estación VMT presenta un patrón diferente al resto de estaciones de monitoreo. Se verifica la existencia de microclimas en los diferentes distritos del AMLC.
3. Los modelos de RLM permitieron reproducir entre un 47 y 88% el pronóstico de las concentraciones horarias de  $PM_{10}$ ,  $PM_{2,5}$ ,  $NO_2$ ,  $O_3$ , CO y  $SO_2$  en la etapa de entrenamiento y, entre 41 a 86%, en la etapa de validación. El primer tipo de modelo implementado para los seis contaminantes obtuvo los mejores indicadores de desempeño, el cual se basó en las concentraciones del mismo contaminante del resto

de estaciones de monitoreo; en comparación con el segundo modelo, el cual se fundamentó en el pronóstico de los contaminantes tomando como predictores el resto de contaminantes y variables meteorológicas de la misma estación de monitoreo. El modelo de pronóstico para el contaminante O<sub>3</sub> obtuvo los mejores indicadores de desempeño a nivel de contaminante, llegando a reproducir el 86% de las concentraciones observadas con una RECM de 5.1 µg/m<sup>3</sup>, en la etapa de validación.

## VI. RECOMENDACIONES

- Se recomienda continuar con la vigilancia de la calidad del aire en el Área Metropolitana de Lima y Callao (AMLC), bajo la administración del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), cuya evaluación es materia prima para la generación de modelos de pronóstico. Así también, garantizar el mantenimiento preventivo y correctivo oportuno de los analizadores automáticos de contaminantes atmosféricos.
- Asimismo, se recomienda validar los resultados obtenidos de la aplicación del análisis estadístico multivariado, de la presente investigación, con el análisis factorial (AF) y el análisis de discriminante lineal (ADL), los cuales son otros dos métodos estadísticos multivariados de gran aplicación en la literatura científica.
- Aplicar técnicas de imputación de datos en base a la inteligencia artificial, como lo son las redes neuronales artificiales, los árboles de decisión y los bosques aleatorios e implementar herramientas de pronóstico basado en técnicas estadísticas y de *machine learning* para la predicción de las concentraciones de los contaminantes atmosféricos en el Área Metropolitana de Lima y Callao, buscando mejorar los indicadores de desempeño de los modelos de pronóstico generados.
- Realizar la evaluación de los patrones espaciales de calidad del aire en el Área Metropolitana de Lima y Callao a nivel estacional, es decir, primavera, verano, otoño e invierno con el fin de determinar la variabilidad temporal de dichos patrones espaciales.
- Investigar la influencia de las variables meteorológicas temperatura de aire, humedad relativa, dirección y velocidad del viento y radiación solar sobre las concentraciones de los contaminantes atmosféricos  $PM_{10}$ ,  $PM_{2,5}$ ,  $NO_2$ ,  $O_3$ ,  $CO$  y  $SO_2$  en el Área Metropolitana de Lima y Callao.

## VII. REFERENCIAS BIBLIOGRÁFICAS

- Aas, W; Mortier, A; Bowersox, V; Cherian, R; Faluvegi, G; Fagerli, H; Myhre, C. 2019. Global and regional trends of atmospheric sulfur. *Scientific reports* 9(1): 1-11.
- Abdullah, S; Ismail, M; Samat, N; Ahmed, A. 2018. Modelling particulate matter (PM<sub>10</sub>) concentration in industrialized area: A comparative study of linear and nonlinear algorithms. *ARPJ. Eng. Appl. Sci* 13: 8227-8235.
- Abdullah, S; Ismail, M; Ahmed, A; Abdullah, A. 2019. Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. *Atmosphere* 10(11): 667.
- Abdullah, S; Nasir, N; Ismail, M; Ahmed, A; Jarkoni, M. 2019. Development of ozone prediction model in urban area.
- Abdullah, S; Napi, N; Ahmed, A; Mansor, W; Mansor, A; Ismail, M; Ramly, Z. 2020. Development of multiple linear regression for particulate matter (PM<sub>10</sub>) forecasting during episodic transboundary haze event in Malaysia. *Atmosphere* 11(3): 289.
- Aldás, J; Uriel, E. 2017. *Análisis multivariante aplicado con R*. 2 ed. Madrid, España, Ediciones Paraninfo. 677 p.
- Al-Karkhi, A; Alqaraghuli, W. (2019). *Applied statistics for environmental science with R*. Elsevier.
- Azid, A; Juahir, H; Ezani, E; Toriman, M; Endut, A; Rahman, M; Umar, R. 2015. Identification source of variation on regional impact of air quality pattern using chemometric. *Aerosol and Air Quality Research* 15(4): 1545-1558.
- Castro, M; Pires, J. 2019. Decision support tool to improve the spatial distribution of air quality monitoring sites. *Atmospheric Pollution Research* 10(3): 827-834.

- Ceylan, Z; Bulkan, S. 2018. Forecasting PM<sub>10</sub> levels using ANN and MLR: A case study for Sakarya City. *Global Nest Journal* 20(2): 281-90
- Corroto, F; Gamarra, O; Barboza, E. 2018. Evaluación multivariante de la calidad del agua en la cuenca del Utcubamba (Perú). *Tecnología y ciencias del agua* 9(5): 33-57.
- DIGESA [Dirección General de Salud Ambiental en Inocuidad Alimentaria]. 2012. II Estudio de saturación de Lima Metropolitana y el Callao 2011 (en Línea). Disponible en [http://www.digesa.minsa.gob.pe/depa/informes\\_tecnicos/Estudio%20de%20Saturacion%202012.pdf](http://www.digesa.minsa.gob.pe/depa/informes_tecnicos/Estudio%20de%20Saturacion%202012.pdf). Consultado 10 enero del 2021.
- DIGESA [Dirección General de Salud Ambiental en Inocuidad Alimentaria]. 2020. Programa Nacional de Vigilancia Sanitaria de Calidad del Aire (en línea). Disponible en [http://www.digesa.minsa.gob.pe/DEPA/aire\\_lc/lima\\_callao.asp](http://www.digesa.minsa.gob.pe/DEPA/aire_lc/lima_callao.asp). Consultado 18 de diciembre 2020.
- Dominick, D; Juahir, H; Latif, M; Zain, S; Aris, A. 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric Environment* 60: 172-181.
- Elbayoumi, M; Ramli, N; Yusof, N. 2015. Development and comparison of regression models and feedforward backpropagation neural network models to predict seasonal indoor PM<sub>2,5-10</sub> and PM<sub>2,5</sub> concentrations in naturally ventilated schools. *Atmospheric Pollution Research* 6(6): 1013-1023.
- Espinoza, J. 2018. Evaluación espacial y temporal del material particulado PM<sub>10</sub> y PM<sub>2,5</sub> en Lima Metropolitana para el periodo 2015-2017. Tesis de grado. Lima, Perú, Universidad Nacional Agraria la Molina. 153 p.
- Espinoza, J; Alderete, M. 2021. Caracterización de regiones espacialmente homogéneas de monóxido de carbono en Lima Metropolitana mediante el algoritmo de clustering k-means. *Biotech & Engineering* 1(1): 17-28.
- Espíritu, C. 2010. Evaluación de la calidad de agua del río Rímac mediante el análisis multivariado. Tesis de Licenciado. Universidad Nacional de Ingeniería.

- Flores, J. 2017. Análisis de una red de monitoreo de calidad de agua en la cuenca del río Jequetepeque mediante análisis multivariado. Tesis de Maestría. Lima, Perú, Pontificia Universidad Católica del Perú. 158 p.
- Flores, R. 2017. Variación temporal de la calidad de agua en la bocatoma “La Atarjea”, Río Rímac (2009-2015). Tesis de Maestría. Lima, Perú, Universidad Nacional Agraria La Molina. 116 p.
- Giordani, P; Brigida, M; Martella, F. 2020. An Introduction to Clustering with R. Singapore, Springer Nature. 340 p.
- Gramsch, E; Cereceda, F; Oyola, P; Baer, D. 2006. Examination of pollution trends in Santiago de Chile with cluster analysis of PM<sub>10</sub> and ozone data. *Atmospheric Environment* 40 (28): 5464-5475.
- Habeebullah, T. 2015. Characterising NO<sub>2</sub>, Its Temporal Variability and Association with Meteorology: A Case Study in Makkah, Saudi Arabia. *Journal of Environment Asia* 8(2): 37-44.
- Ibáñez, L. 2008. Comparación entre las metodologías de muestreo equivalentes y de referencia del material particulado atmosférico, en la ciudad de Lima. Tesis de Grado en Ing. Ambiental. Lima, Perú, Universidad Nacional Agraria la Molina.
- Ibarra-Berastegi, G; Sáenz, J; Ezcurra, A; Ganzedo, U; de Argandoña, J; Errasti, I; Polanco-Martínez, J. 2009. Assessing spatial variability of SO<sub>2</sub> field as detected by an air quality network using self-organizing maps, cluster, and principal component analysis. *Atmospheric Environment* 43(25): 3829-3836.
- INECC (Instituto Nacional de Ecología y Cambio Climático). 2018. Evaluación de redundancia de las estaciones de monitoreo de calidad del aire que operan en el Estado de México. Coordinación General de Contaminación y Salud Ambiental, Dirección de Investigación sobre la Calidad del Aire y los Contaminantes Climáticos. Ciudad de México. Noviembre 2018.
- Isiyaka, H; Azid, A. 2015. Air quality pattern assessment in Malaysia using multivariate techniques. *Malaysian Journal of Analytical Sciences* 19(5): 966-978.



- Kachigan, S. K. 1991. *Multivariate statistical analysis: A conceptual introduction*. Radius Press.
- Karl, W; Hlávka, Z. 2015. *Multivariate Statistics: Exercises and Solutions*. 2 ed. Berlin, Springer. 362 p.
- Karl, W; Simar, L. 2019. *Applied Multivariate Statistical Analysis*. 5 ed. Switzerland, Springer Nature. 558 p.
- Kim, K; Kabir, E; Kabir, S. 2015. A review on the human health impact of airborne particulate matter. *Environment international* 74:136-143.
- Lu, W; He, H; Dong, L. 2011. Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Building and Environment* 46 (3): 577-583.
- Malley, C; Schneidmesser, E; Moller, S; Hicks, W; Braban, C; Heal, M. 2018. Analysis of the distributions of hourly NO<sub>2</sub> concentrations contributing to annual average NO<sub>2</sub> concentrations across the European monitoring network between 2000 and 2014. *Atmospheric Chemistry and Physics*: 3563-3587.
- MINAM [Ministerio del Ambiente]. 2005. Ley N° 28611: Ley General del Ambiente. Lima, Perú.
- MINAM [Ministerio del Ambiente]. 2016. Resolución Ministerial N° 181-2016-MINAM: Establecen el índice de Calidad del Aire – INCA y crean el Sistema de Información de Calidad del Aire – INFO AIRE PERÚ. Lima, Perú.
- MINAM [Ministerio del Ambiente]. 2017. Decreto Supremo N° 003-2017-MINAM: Aprueban Estándares de Calidad Ambiental (ECA) para Aire y establecen disposiciones complementarias. Lima, Perú.
- MINAM [Ministerio del Ambiente]. 2019. Decreto Supremo N° 010-2019-MINAM: Decreto Supremo que aprueba el Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire. Lima, Perú.

- MINAM [Ministerio del Ambiente]. 2019. Diagnóstico de la Gestión de la Calidad del Aire de Lima y Callao. Informe N° 00283-2019-MINAM/VMGA/DGCA/DCAE. Lima, Perú.
- MINSA [Ministerio de Salud]. 2003. Decreto Supremo N° 009-2003-SA. Reglamento de los Niveles de Estados de Alerta Nacionales para Contaminantes del Aire. Lima, Perú.
- MINSA [Ministerio de Salud]. 2005. Decreto Supremo N° 012-2005-SA. Modifican Reglamento de los Niveles de Estados de Alerta Nacionales para Contaminantes del Aire. Lima, Perú.
- Moura, I; Barreto, A; Cesar, R; Cruz, A; Menezes, M. 2018. Use of statistical multivariate analysis to identify sources emitting particulate matter in Belo Horizonte, Brazil. *Revista Ambiente & Água* 13(5).
- Munir, S; Habeebullah, T; Mohammed, A; Morsy, E; Awad, A; Seroji, A; Hassan, I. 2015. An Analysis into the Temporal Variations of Ground Level Ozone in the Arid Climate of Makkah applying k-means Algorithms. *Environment Asia* 8(1).
- Napi, N; Mohamed, M; Abdullah, S; Mansor, A; Ahmed, A; Ismail, M. 2020. Multiple Linear Regression (MLR) and Principal Component Regression (PCR) for Ozone (O<sub>3</sub>) Concentrations Prediction. In *IOP Conference Series: Earth and Environmental Science* (Vol. 616, No. 1, p. 012004). IOP Publishing.
- Núñez, D; Pérez, L; Manzoor, S; Cáceres, J. 2019. Statistical tools for air pollution assessment: multivariate and spatial analysis studies in the Madrid region. *Journal of Analytical Methods in Chemistry* 2019: 1-9.
- Ordieres, J; Vergara, E; Capuz, R; Salazar, R. 2005. Neural network prediction model for fine particulate matter (PM<sub>2.5</sub>) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling & Software* 20(5): 547-559.
- Pacsi, S. 2016. Análisis temporal y espacial de la calidad del aire determinado por material particulado PM<sub>10</sub> y PM<sub>2,5</sub> en Lima Metropolitana. *Anales Científicos* 77 (2): 273-283.

- Pires, J; Martins, F; Sousa, S; Alvim-Ferraz, M; Pereira, M. 2008. Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling & Software* 23(1): 50-55.
- Pires, J; Sousa, S; Pereira, M; Alvim-Ferraz, M; Martins, F. 2008. Management of air quality monitoring using principal component and cluster analysis—Part I: SO<sub>2</sub> and PM<sub>10</sub>. *Atmospheric Environment* 42 (6): 1249-1260.
- Pires, J; Sousa, S; Pereira, M; Alvim-Ferraz, M; Martins, F. 2008. Management of air quality monitoring using principal component and cluster analysis—Part II: CO, NO<sub>2</sub> and O<sub>3</sub>. *Atmospheric Environment* 42 (6): 1261-1274.
- Pires, J; Martins, F; Alvim-Ferraz, M; Pereira, M. 2009. Evaluation of redundant measurements on the air quality monitoring network of Lisbon and Tagus Valley. *Chemical Product and Process Modeling* 4(4).
- Pires, J; Pereira, M; Alvim-Ferraz, M; Martins, F. 2009. Identification of redundant air quality measurements through the use of principal component analysis. *Atmospheric environment* 43(25): 3837-3842.
- Polanco, J. 2016. El papel del análisis por componentes principales en la evaluación de redes de control de la calidad del aire. *Comunicaciones en Estadística* 9 (2): 271-294.
- Rahimi, A. 2017. Short-term prediction of NO<sub>2</sub> and NO<sub>x</sub> concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran. *Ecological Processes* 6(1): 1-9.
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Rozante, J; Rozante, V; Souza Alvim, D., Ocimar, A; Barboza J; Siqueira M; Moreira, D. 2017. Variations of carbon monoxide concentrations in the megacity of São Paulo from 2000 to 2015 in different time scales. *Atmosphere* 8(5): 81.
- RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

- Sánchez, O; Ordóñez, C; Aliaga, R; Rojas, J. 2015. Evaluación de la Calidad del Aire en Lima Metropolitana, 2014. Lima: Servicio Nacional de Meteorología e Hidrología del Perú.
- Sánchez, O; Ordóñez, C. 2016. Evaluación de la calidad del aire en Lima Metropolitana 2015. Lima: SENAMHI.
- SENAMHI (Servicio Nacional de Meteorología e Hidrología del Perú). 2016. Gestión del Monitoreo de la Calidad del Aire: Gases y Material Particulado en Lima. Lima, Perú.
- Shah, M; Shaheen, N. 2008. Annual and seasonal variations of trace metals in atmospheric suspended particulate matter in Islamabad, Pakistan. *Water Air Soil Pollut* 190:13-25.
- Shams, S; Jahani, A; Moeinaddini, M; Khorasani, N. 2020. Air carbon monoxide forecasting using an artificial neural network in comparison with multiple regression. *Modeling Earth Systems and Environment* 6(3): 1467-1475.
- Silva, J; Rojas, J; Norabuena, M; Molina, C; Toro, R; Leiva, M. 2017. Particulate matter levels in a South American megacity: the metropolitan area of Lima-Callao, Peru. *Environmental Monitoring and Assessment* 189 (12): 635.
- Silva, J; Rojas, J; Norabuena, M; Seguel, R. 2018. Ozone and volatile organic compounds in the metropolitan area of Lima-Callao, Peru. *Air Quality, Atmosphere & Health* 11(8): 993-1008.
- Sousa, S; Martins, F; Alvim-Ferraz, M; Pereira, M. 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software* 22 (1): 97-103.
- Srimuruganandam, B; Nagendra, S. 2012. Source characterization of PM<sub>10</sub> and PM<sub>2.5</sub> mass using a chemical mass balance model at urban roadside. *Science of the Total Environment* 433:8-19.
- Ul-Saufie, A; Yahaya, A; Ramli, N; Rosaida, N; Hamid, H. 2013. Future daily PM<sub>10</sub> concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmospheric Environment* 77: 621-630.

- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2019. Sulfur Dioxide (SO<sub>2</sub>) Pollution. Disponible en: <https://www.epa.gov/so2-pollution>.
- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2019. Nitrogen Dioxide (NO<sub>2</sub>) Pollution. Disponible en: <https://www.epa.gov/no2-pollution>.
- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2019. Carbon Monoxide (CO) Pollution in Outdoor Air. Disponible en: <https://www.epa.gov/co-pollution>.
- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2020. Ground-level Ozone Pollution. Disponible en: <https://www.epa.gov/ground-level-ozone-pollution>
- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2020. Particulate Matter (PM) Pollution. Disponible en: <https://www.epa.gov/pm-pollution>.
- USEPA (Agencia de Protección Ambiental de Estados Unidos). 2020. La contaminación atmosférica. Disponible en: <https://espanol.epa.gov/espanol/aire#contaminacion>.
- Wark, K; Warner, C. 2007. Contaminación del aire. Origen y control. México, D.F.: Limusa Noriega Editores.
- Wehrens, R. 2020. Chemometrics with R: Multivariate Data Analysis in the Natural and Life Sciences. 2 ed. Germany, Springer Nature. 308 p.
- Wang, C; Zhao, L; Sun, W; Xue, J; Xie, Y. 2018. Identifying redundant monitoring stations in an air quality monitoring network. Atmospheric Environment 190: 256-268.
- WHO (Organización Mundial de la Salud). 2016. WHO Global Urban Ambient Air Pollution Database (update 2016). Disponible en: [https://www.who.int/phe/health\\_topics/outdoorair/databases/cities/en/](https://www.who.int/phe/health_topics/outdoorair/databases/cities/en/)
- WHO (Organización Mundial de la Salud). 2018. Calidad del aire y salud: Datos y cifras. Disponible en: [https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- WHO (Organización Mundial de la Salud). 2018. Nueve de cada diez personas de todo el mundo respiran aire contaminado: Sin embargo, cada vez hay más países que toman

medidas. Disponible en: <https://www.who.int/es/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action>

Zamora, A; Quesada, C; Hurtado, L; Mondéjar, D. 2020. An Introduction to Data Analysis in R: Hands-on Coding, Data Mining, Visualization and Statistics from Scratch. Switzerland, Springer Nature. 276 p.

Zelterman, D. 2015. Applied Multivariate Statistics with R: Statistics for Biology and Health. Switzerland, Springer. 393 p.

Zhang, J; Ding, W. 2017. Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. International journal of environmental research and public health 14(2): 114.

Zhao, L; Xie, Y; Wang, J; Xu, X. 2015. A performance assessment and adjustment program for air quality monitoring networks in Shanghai. Atmospheric Environment 122: 382-392.

## VIII. ANEXOS

### Anexo 1: Salidas del modelo I para el pronóstico de las concentraciones horarias de contaminantes atmosféricos para la estación ATE

#### Entrenamiento PM<sub>10</sub>

Call:

```
lm(formula = ATE ~ CDM + STA + VMT + HCH + SMP, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-87.14	-19.89	-3.16	15.55	143.32

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	19.26696	1.70214	11.319	< 2e-16	***
CDM	-0.19331	0.04985	-3.878	0.000108	***
STA	0.77630	0.02693	28.828	< 2e-16	***
VMT	-0.03614	0.00769	-4.701	2.71e-06	***
HCH	0.42057	0.01919	21.920	< 2e-16	***
SMP	0.16873	0.03170	5.322	1.10e-07	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.82 on 3093 degrees of freedom

Multiple R-squared: 0.5969, Adjusted R-squared: 0.5962

F-statistic: 915.9 on 5 and 3093 DF, p-value: < 2.2e-16

#### Validación PM<sub>10</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm10_ATE)
```

Residuals:

Min	1Q	Median	3Q	Max
-73.569	-13.906	-0.567	13.433	77.827

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	47.15422	2.16940	21.74	<2e-16	***
observado	0.58642	0.01746	33.58	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.27 on 773 degrees of freedom

Multiple R-squared: 0.5933, Adjusted R-squared: 0.5928

F-statistic: 1128 on 1 and 773 DF, p-value: < 2.2e-16

### Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = ATE ~ STA + HCH + SJL + PPD, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-33.819	-6.274	-0.416	5.446	34.585

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.41043	1.03332	4.268	2.15e-05	***
STA	0.40384	0.03811	10.597	< 2e-16	***
HCH	0.29885	0.02845	10.506	< 2e-16	***
SJL	0.20051	0.03500	5.729	1.33e-08	***
PPD	0.10379	0.03351	3.098	0.002	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.04 on 1020 degrees of freedom

Multiple R-squared: 0.5796, Adjusted R-squared: 0.578

F-statistic: 351.6 on 4 and 1020 DF, p-value: < 2.2e-16

### Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_ATE)
```

Residuals:

Min	1Q	Median	3Q	Max
-19.331	-5.742	-0.528	4.686	32.668

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	17.93740	1.26672	14.16	<2e-16	***
observado	0.48356	0.03333	14.51	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.177 on 254 degrees of freedom

Multiple R-squared: 0.4532, Adjusted R-squared: 0.451

F-statistic: 210.5 on 1 and 254 DF, p-value: < 2.2e-16

### Entrenamiento NO<sub>2</sub>

Call:

```
lm(formula = ATE ~ SBJ + CDM + STA + VMT + HCH + SJL + PPD, data = training)
```

Residuals:



	Min	1Q	Median	3Q	Max
	-50.181	-10.928	-1.499	8.626	72.595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	9.52897	1.60663	5.931	3.59e-09	***
SBJ	0.25640	0.02539	10.097	< 2e-16	***
CDM	-0.57567	0.05006	-11.500	< 2e-16	***
STA	-0.31824	0.03851	-8.264	2.68e-16	***
VMT	1.50615	0.05461	27.580	< 2e-16	***
HCH	0.20952	0.04989	4.199	2.80e-05	***
SJL	0.42540	0.05279	8.058	1.38e-15	***
PPD	-0.09452	0.02838	-3.330	0.000886	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.35 on 1828 degrees of freedom  
 Multiple R-squared: 0.5969, Adjusted R-squared: 0.5953  
 F-statistic: 386.7 on 7 and 1828 DF, p-value: < 2.2e-16

### Validación NO<sub>2</sub>

Call:

lm(formula = pronosticado ~ observado, data = no2\_ATE)

Residuals:

	Min	1Q	Median	3Q	Max
	-35.738	-8.396	-2.166	6.512	46.721

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	16.02055	1.07072	14.96	<2e-16	***
observado	0.59208	0.02414	24.53	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.73 on 457 degrees of freedom  
 Multiple R-squared: 0.5684, Adjusted R-squared: 0.5674  
 F-statistic: 601.8 on 1 and 457 DF, p-value: < 2.2e-16

### Entrenamiento O<sub>3</sub>

Call:

lm(formula = ATE ~ STA + VMT + HCH + SJL + CRB + PPD, data = training)

Residuals:

	Min	1Q	Median	3Q	Max
	-17.5013	-1.4366	0.0433	1.2247	26.3534

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.07604	0.22058	-0.345	0.7304	
STA	0.35319	0.02722	12.977	< 2e-16	***

VMT	-0.04129	0.01485	-2.781	0.0055	**
HCH	0.21892	0.01967	11.132	< 2e-16	***
SJL	0.16448	0.02252	7.305	5.03e-13	***
CRB	-0.11474	0.01188	-9.654	< 2e-16	***
PPD	0.23833	0.01531	15.569	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.193 on 1206 degrees of freedom  
Multiple R-squared: 0.8769, Adjusted R-squared: 0.8762  
F-statistic: 1431 on 6 and 1206 DF, p-value: < 2.2e-16

### Validación O<sub>3</sub>

Call:

lm(formula = pronosticado ~ observado, data = o3\_ATE)

Residuals:

	Min	1Q	Median	3Q	Max
	-24.0712	-1.5659	-0.2972	1.5009	12.6129

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.33368	0.32986	4.043	6.7e-05	***
observado	0.81671	0.01882	43.389	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.372 on 301 degrees of freedom  
Multiple R-squared: 0.8622, Adjusted R-squared: 0.8617  
F-statistic: 1883 on 1 and 301 DF, p-value: < 2.2e-16

### Entrenamiento CO

Call:

lm(formula = ATE ~ ., data = training)

Residuals:

	Min	1Q	Median	3Q	Max
	-1511.52	-241.93	19.68	272.50	1443.76

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	197.37157	39.65726	4.977	6.82e-07	***
SBJ	0.10295	0.03724	2.765	0.00573	**
CDM	-0.48212	0.06843	-7.046	2.28e-12	***
STA	0.86442	0.03489	24.773	< 2e-16	***
VMT	0.50292	0.04955	10.149	< 2e-16	***
HCH	0.53671	0.03500	15.336	< 2e-16	***
SJL	-0.36108	0.03386	-10.664	< 2e-16	***
SMP	-0.23637	0.03256	-7.259	4.96e-13	***
CRB	-0.20314	0.03374	-6.020	1.95e-09	***
PPD	0.46879	0.02004	23.397	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 431.4 on 2997 degrees of freedom  
Multiple R-squared: 0.6002, Adjusted R-squared: 0.599  
F-statistic: 499.8 on 9 and 2997 DF, p-value: < 2.2e-16

### Validación CO

Call:  
lm(formula = pronosticado ~ observado, data = co\_ATE)

Residuals:

Min	1Q	Median	3Q	Max
-1037.97	-231.68	-14.94	233.21	1189.28

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	744.53419	33.96457	21.92	<2e-16 ***
observado	0.58713	0.01858	31.60	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 335.2 on 750 degrees of freedom  
Multiple R-squared: 0.5711, Adjusted R-squared: 0.5706  
F-statistic: 998.8 on 1 and 750 DF, p-value: < 2.2e-16

### Entrenamiento SO<sub>2</sub>

Call:  
lm(formula = ATE ~ SBJ + CDM + STA + SJL + SMP + CRB + PPD, data = training)

Residuals:

Min	1Q	Median	3Q	Max
-29.119	-4.774	-0.808	4.056	41.052

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.59885	2.16067	4.443	1.09e-05 ***
SBJ	1.14451	0.19846	5.767	1.40e-08 ***
CDM	-2.97728	0.39223	-7.591	1.51e-13 ***
STA	0.75161	0.07057	10.650	< 2e-16 ***
SJL	0.88918	0.16160	5.502	5.92e-08 ***
SMP	1.99727	0.25299	7.895	1.78e-14 ***
CRB	-1.33009	0.25448	-5.227	2.52e-07 ***
PPD	0.21233	0.07285	2.915	0.00372 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.629 on 513 degrees of freedom  
Multiple R-squared: 0.6305, Adjusted R-squared: 0.6254  
F-statistic: 125 on 7 and 513 DF, p-value: < 2.2e-16

## Validación SO<sub>2</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = so2_ATE)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-20.429	-5.584	-1.313	5.175	24.781

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	12.60325	1.92028	6.563	1.19e-09	***
observado	0.58720	0.05584	10.515	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.162 on 128 degrees of freedom

Multiple R-squared: 0.4635, Adjusted R-squared: 0.4593

F-statistic: 110.6 on 1 and 128 DF, p-value: < 2.2e-16

## Anexo 2: Salidas del modelo II para el pronóstico de las concentraciones horarias de contaminantes atmosféricos para la estación ATE

### Entrenamiento PM<sub>10</sub>

```
Call:
lm(formula = pm10 ~ pm25 + co + no2 + o3 + temp + HR + ws + wd +
    rad, data = training)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-76.609 -20.287  -2.056   17.823   95.785
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 231.518825  21.185857  10.928 < 2e-16 ***
pm25         0.813434   0.038751  20.991 < 2e-16 ***
co           0.021346   0.001792  11.912 < 2e-16 ***
no2          0.475923   0.047603   9.998 < 2e-16 ***
o3           -1.405345   0.094194 -14.920 < 2e-16 ***
temp        -3.333419   0.434752  -7.667 2.67e-14 ***
HR           -1.294101   0.153844  -8.412 < 2e-16 ***
ws           -5.002774   1.530962  -3.268 0.0011 **
wd           -0.053456   0.024984  -2.140 0.0325 *
rad           0.047225   0.004554  10.370 < 2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 28.27 on 2094 degrees of freedom
Multiple R-squared:  0.4936, Adjusted R-squared:  0.4915
F-statistic: 226.8 on 9 and 2094 DF,  p-value: < 2.2e-16
```

### Validación PM<sub>10</sub>

```
Call:
lm(formula = pronosticado ~ observado, data = pm10_ATE)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-45.046 -14.193  -1.639   12.558   67.432
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.26381    2.75279   21.16 <2e-16 ***
observado    0.49610    0.02239   22.16 <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 19.75 on 524 degrees of freedom
Multiple R-squared:  0.4837, Adjusted R-squared:  0.4827
F-statistic: 491 on 1 and 524 DF,  p-value: < 2.2e-16
```

### Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = pm25 ~ pm10 + co + no2 + so2 + o3 + temp + HR +  
    ws, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-68.066	-9.207	-1.411	8.511	61.897

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-7.858e+01	1.037e+01	-7.581	5.12e-14	***
pm10	1.988e-01	9.825e-03	20.233	< 2e-16	***
co	2.770e-03	9.125e-04	3.035	0.00243	**
no2	3.051e-01	2.454e-02	12.432	< 2e-16	***
so2	2.405e-01	2.669e-02	9.010	< 2e-16	***
o3	3.708e-01	3.493e-02	10.616	< 2e-16	***
temp	5.071e-01	2.328e-01	2.179	0.02947	*
HR	7.860e-01	7.716e-02	10.187	< 2e-16	***
ws	-4.401e+00	7.523e-01	-5.850	5.70e-09	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.22 on 2095 degrees of freedom

Multiple R-squared: 0.479, Adjusted R-squared: 0.477

F-statistic: 240.8 on 8 and 2095 DF, p-value: < 2.2e-16

### Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_ATE)
```

Residuals:

Min	1Q	Median	3Q	Max
-30.970	-6.616	-0.339	6.164	35.074

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	21.86071	0.94395	23.16	<2e-16	***
observado	0.42645	0.02215	19.25	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.18 on 524 degrees of freedom

Multiple R-squared: 0.4144, Adjusted R-squared: 0.4133

F-statistic: 370.8 on 1 and 524 DF, p-value: < 2.2e-16

### Entrenamiento NO<sub>2</sub>

Call:

```
lm(formula = no2 ~ pm10 + pm25 + co + so2 + o3 + temp + HR +  
    ws + rad, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-31.289	-7.982	-1.456	6.337	44.191

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-6.437473	8.853320	-0.727	0.46723
pm10	0.091162	0.009038	10.086	< 2e-16 ***
pm25	0.220754	0.017761	12.429	< 2e-16 ***
co	0.011680	0.000756	15.451	< 2e-16 ***
so2	-0.350514	0.021939	-15.977	< 2e-16 ***
o3	-0.463419	0.042143	-10.996	< 2e-16 ***
temp	2.013049	0.191351	10.520	< 2e-16 ***
HR	-0.260954	0.066348	-3.933	8.66e-05 ***
ws	2.057389	0.654332	3.144	0.00169 **
rad	0.018346	0.001967	9.328	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.98 on 2094 degrees of freedom

Multiple R-squared: 0.5199, Adjusted R-squared: 0.5178

F-statistic: 252 on 9 and 2094 DF, p-value: < 2.2e-16

## Validación NO<sub>2</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = no2_ATE)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.0049	-6.2968	0.0088	5.1826	27.2575

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	19.59565	0.90526	21.65	<2e-16 ***
observado	0.47038	0.02168	21.70	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.761 on 524 degrees of freedom

Multiple R-squared: 0.4732, Adjusted R-squared: 0.4722

F-statistic: 470.7 on 1 and 524 DF, p-value: < 2.2e-16

## Entrenamiento O<sub>3</sub>

Call:

```
lm(formula = o3 ~ ., data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-25.4459	-3.2098	-0.2483	3.1769	29.8628

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
--	----------	------------	---------	----------

```

(Intercept) 63.4586571 4.4323803 14.317 < 2e-16 ***
pm10        -0.0616545 0.0044639 -13.812 < 2e-16 ***
pm25         0.0752396 0.0091156  8.254 2.68e-16 ***
co           0.0068041 0.0003731 18.235 < 2e-16 ***
no2         -0.1175238 0.0106921 -10.992 < 2e-16 ***
so2         -0.1380641 0.0113390 -12.176 < 2e-16 ***
temp        -0.9249076 0.0967661  -9.558 < 2e-16 ***
HR          -0.4401223 0.0320975 -13.712 < 2e-16 ***
ws           2.1812940 0.3306620  6.597 5.31e-11 ***
wd          -0.0163490 0.0053392  -3.062 0.00223 **
rad          0.0330481 0.0007060 46.813 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 6.027 on 2093 degrees of freedom  
Multiple R-squared: 0.8039, Adjusted R-squared: 0.803  
F-statistic: 858.1 on 10 and 2093 DF, p-value: < 2.2e-16

### Validación O<sub>3</sub>

```

Call:
lm(formula = pronosticado ~ observado, data = o3_ATE)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-14.6582  -3.0352  -0.3714   2.1674  28.2421

```

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.19783    0.30986   7.093 4.28e-12 ***
observado    0.82738    0.01799 45.984 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 5.124 on 524 degrees of freedom  
Multiple R-squared: 0.8014, Adjusted R-squared: 0.801  
F-statistic: 2115 on 1 and 524 DF, p-value: < 2.2e-16

### Entrenamiento CO

```

Call:
lm(formula = co ~ pm10 + pm25 + no2 + so2 + o3 + temp + HR +
    rad, data = training)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-1373.62  -213.44   15.08   219.40  1126.05

```

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -334.09353  236.47487  -1.413  0.1579
pm10         2.76384    0.24549  11.258 < 2e-16 ***
pm25         1.28702    0.49971   2.576  0.0101 *
no2          8.77856    0.56535  15.528 < 2e-16 ***

```



```

so2          5.42325    0.61321    8.844 < 2e-16 ***
o3           20.11840    1.09090   18.442 < 2e-16 ***
temp        -21.49782    5.35093   -4.018 6.09e-05 ***
HR           14.71353    1.76404    8.341 < 2e-16 ***
rad          -0.65298    0.05158  -12.658 < 2e-16 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 328 on 2095 degrees of freedom  
Multiple R-squared: 0.4742, Adjusted R-squared: 0.4722  
F-statistic: 236.2 on 8 and 2095 DF, p-value: < 2.2e-16

### Validación CO

Call:  
lm(formula = pronosticado ~ observado, data = co\_ATE)

```

Residuals:
    Min       1Q   Median       3Q      Max
-838.37 -126.39   -6.86  141.27  579.33

```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  730.5373    29.6630   24.63 <2e-16 ***
observado     0.4741     0.0203   23.35 <2e-16 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 213 on 524 degrees of freedom  
Multiple R-squared: 0.5099, Adjusted R-squared: 0.509  
F-statistic: 545.2 on 1 and 524 DF, p-value: < 2.2e-16

### Entrenamiento SO<sub>2</sub>

Call:  
lm(formula = so2 ~ pm25 + co + no2 + o3 + temp + HR + ws + wd + rad, data = training)

```

Residuals:
    Min       1Q   Median       3Q      Max
-35.988  -5.873    0.741    6.740   42.275

```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.8325254    8.4190740  -0.218 0.827712
pm25         0.1648560    0.0153993   10.705 < 2e-16 ***
co           0.0066549    0.0007121    9.345 < 2e-16 ***
no2         -0.2996860    0.0189171  -15.842 < 2e-16 ***
o3          -0.5014674    0.0374319  -13.397 < 2e-16 ***
temp        2.5594447    0.1727668   14.814 < 2e-16 ***
HR          -0.2981585    0.0611362   -4.877 1.16e-06 ***
ws          -5.8134953    0.6083907   -9.556 < 2e-16 ***
wd           0.0371979    0.0099285    3.747 0.000184 ***
rad         0.0152830    0.0018097    8.445 < 2e-16 ***

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.23 on 2094 degrees of freedom

Multiple R-squared: 0.5079, Adjusted R-squared: 0.5058

F-statistic: 240.1 on 9 and 2094 DF, p-value: < 2.2e-16

### Validación SO<sub>2</sub>

Call:

lm(formula = pronosticado ~ observado, data = so2\_ATE)

Residuals:

Min	1Q	Median	3Q	Max
-21.4470	-5.6124	-0.2575	4.7783	23.5863

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	13.09110	0.68372	19.15	<2e-16	***
observado	0.51755	0.02126	24.34	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.029 on 524 degrees of freedom

Multiple R-squared: 0.5307, Adjusted R-squared: 0.5299

F-statistic: 592.7 on 1 and 524 DF, p-value: < 2.2e-16

### Anexo 3: Salidas de los modelos I y II para el pronóstico de las concentraciones horarias de PM<sub>10</sub> y PM<sub>2,5</sub> para las estaciones VMT, CDM y PPD

#### Estación VMT – Modelo I - Entrenamiento PM<sub>10</sub>

Call:

```
lm(formula = VMT ~ ., data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-163.304	-41.275	-5.517	35.542	227.209

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-25.51404	3.94850	-6.462	1.20e-10	***
ATE	-0.14329	0.03741	-3.830	0.000131	***
SBJ	0.55868	0.08546	6.538	7.29e-11	***
CDM	-0.24683	0.11788	-2.094	0.036343	*
STA	0.53574	0.06748	7.940	2.81e-15	***
HCH	0.19563	0.04824	4.055	5.13e-05	***
SJL	-0.49383	0.06027	-8.194	3.67e-16	***
SMP	0.36887	0.06696	5.509	3.90e-08	***
CRB	0.69943	0.04357	16.055	< 2e-16	***
PPD	0.49590	0.02927	16.942	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.07 on 3089 degrees of freedom

Multiple R-squared: 0.3988, Adjusted R-squared: 0.3971

F-statistic: 227.7 on 9 and 3089 DF, p-value: < 2.2e-16

#### Estación VMT – Modelo I - Validación PM<sub>10</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm10_VMT)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-124.228	-25.878	-2.113	23.841	138.123

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	74.89435	2.58394	28.98	<2e-16	***
observado	0.37380	0.01765	21.18	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.91 on 773 degrees of freedom

Multiple R-squared: 0.3673, Adjusted R-squared: 0.3665

F-statistic: 448.7 on 1 and 773 DF, p-value: < 2.2e-16

#### Estación VMT – Modelo II - Entrenamiento PM<sub>10</sub>

Call:

```
lm(formula = pm10 ~ pm25 + co + no2 + so2 + o3 + temp + HR +
```

```
wd + rad, data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-186.986	-32.495	-5.756	27.894	215.135

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	247.058855	18.074034	13.669	< 2e-16	***
pm25	1.850044	0.062723	29.496	< 2e-16	***
co	0.079269	0.002637	30.060	< 2e-16	***
no2	1.860440	0.085797	21.684	< 2e-16	***
so2	1.260211	0.203015	6.207	5.77e-10	***
o3	-1.082022	0.091036	-11.886	< 2e-16	***
temp	0.887853	0.376774	2.356	0.0185	*
HR	-2.799311	0.129852	-21.558	< 2e-16	***
wd	-0.095315	0.012075	-7.893	3.51e-15	***
rad	-0.024732	0.004305	-5.745	9.67e-09	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49.53 on 5592 degrees of freedom

Multiple R-squared: 0.5979, Adjusted R-squared: 0.5972

F-statistic: 923.8 on 9 and 5592 DF, p-value: < 2.2e-16

### Estación VMT – Modelo II - Validación PM<sub>10</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm10_VMT)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-122.920	-27.205	-1.321	25.596	150.435

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	55.25550	2.03006	27.22	<2e-16	***
observado	0.59313	0.01309	45.32	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.29 on 1398 degrees of freedom

Multiple R-squared: 0.595, Adjusted R-squared: 0.5948

F-statistic: 2054 on 1 and 1398 DF, p-value: < 2.2e-16

### Estación CDM – Modelo I - Entrenamiento PM<sub>10</sub>

Call:

```
lm(formula = CDM ~ ATE + SBJ + STA + VMT + SJL + SMP + CRB +  
PPD, data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-29.209	-6.064	-0.429	5.504	44.321

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.433755	0.601727	4.045	5.37e-05	***
ATE	-0.022014	0.005402	-4.075	4.71e-05	***
SBJ	0.316777	0.011796	26.855	< 2e-16	***
STA	0.088551	0.010226	8.660	< 2e-16	***
VMT	-0.005991	0.002735	-2.190	0.0286	*
SJL	0.073726	0.008480	8.694	< 2e-16	***
SMP	0.165710	0.009788	16.929	< 2e-16	***
CRB	-0.063170	0.006785	-9.310	< 2e-16	***
PPD	0.031921	0.004628	6.898	6.38e-12	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.163 on 3090 degrees of freedom

Multiple R-squared: 0.5766, Adjusted R-squared: 0.5755

F-statistic: 526 on 8 and 3090 DF, p-value: < 2.2e-16

### Estación CDM – Modelo I - Validación PM<sub>10</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm10_CDM)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-17.8590	-4.9018	-0.5924	4.7097	24.2399

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	14.8446	0.6549	22.67	<2e-16	***
observado	0.5470	0.0178	30.73	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.152 on 773 degrees of freedom

Multiple R-squared: 0.5498, Adjusted R-squared: 0.5492

F-statistic: 944 on 1 and 773 DF, p-value: < 2.2e-16

### Estación CDM – Modelo II - Entrenamiento PM<sub>10</sub>

Call:

```
lm(formula = pm10 ~ pm25 + co + no2 + so2 + temp + HR + ws + wd, data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-35.806	-6.532	-0.636	6.008	42.583

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	67.026970	4.292458	15.615	< 2e-16	***

```

pm25      1.226505   0.031575   38.844 < 2e-16 ***
co        0.016968   0.001305   13.000 < 2e-16 ***
no2       0.438962   0.026298   16.692 < 2e-16 ***
so2       0.706287   0.064168   11.007 < 2e-16 ***
temp     -1.340943   0.057613  -23.275 < 2e-16 ***
HR       -0.477972   0.028896  -16.541 < 2e-16 ***
ws       -0.802351   0.170180   -4.715 2.49e-06 ***
wd        0.024779   0.007987    3.103 0.00193 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.04 on 4549 degrees of freedom
Multiple R-squared:  0.5804, Adjusted R-squared:  0.5797

F-statistic: 786.6 on 8 and 4549 DF,  p-value: < 2.2e-16

```

### Estación CDM – Modelo II - Validación PM<sub>10</sub>

```

Call:
lm(formula = pronosticado ~ observado, data = pm10_CDM)

Residuals:
    Min       1Q   Median       3Q      Max
-25.4839  -5.1430  -0.4797   5.1613  25.6269

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.56092   0.57766   28.67 <2e-16 ***
observado    0.55210   0.01458   37.88 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.694 on 1138 degrees of freedom
Multiple R-squared:  0.5577, Adjusted R-squared:  0.5573

F-statistic: 1435 on 1 and 1138 DF,  p-value: < 2.2e-16

```

### Estación PPD – Modelo I - Entrenamiento PM<sub>10</sub>

```

Call:
lm(formula = PPD ~ CDM + VMT + SJL + SMP + CRB, data = training)

Residuals:
    Min       1Q   Median       3Q      Max
-102.285  -24.901  -4.217   20.678  123.711

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.264468   2.167487  13.502 < 2e-16 ***
CDM          0.525383   0.058090   9.044 < 2e-16 ***
VMT          0.173630   0.009778  17.758 < 2e-16 ***
SJL          0.401431   0.027963  14.356 < 2e-16 ***
SMP         -0.090191   0.039229  -2.299 0.02157 *
CRB          0.070247   0.026011   2.701 0.00696 **

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.35 on 3093 degrees of freedom
Multiple R-squared:  0.3676, Adjusted R-squared:  0.3666

F-statistic: 359.5 on 5 and 3093 DF,  p-value: < 2.2e-16

```

### Estación PPD – Modelo I - Validación PM<sub>10</sub>

```

Call:
lm(formula = pronosticado ~ observado, data = pm10_PPD)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-67.166 -16.211  -0.648  13.008  79.221

```

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 66.49949   2.04501   32.52  <2e-16 ***
observado    0.36948   0.01811   20.40  <2e-16 ***

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.95 on 773 degrees of freedom
Multiple R-squared:  0.3499, Adjusted R-squared:  0.349

F-statistic:  416 on 1 and 773 DF,  p-value: < 2.2e-16

```

### Estación PPD – Modelo II - Entrenamiento PM<sub>10</sub>

```

Call:
lm(formula = pm10 ~ ., data = training)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-92.576 -21.369  -2.043  18.767 123.075

```

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.755e+02  9.396e+00  29.323  < 2e-16 ***
pm25        1.675e+00  3.257e-02  51.413  < 2e-16 ***
co          1.346e-02  8.864e-04  15.183  < 2e-16 ***
no2         2.391e-01  2.593e-02   9.222  < 2e-16 ***
so2        -1.457e-01  3.524e-02  -4.134  3.60e-05 ***
o3         -5.682e-01  4.613e-02 -12.317  < 2e-16 ***
temp       -7.504e-01  1.938e-01  -3.872  0.000109 ***
HR         -2.454e+00  7.118e-02 -34.470  < 2e-16 ***
ws          3.567e+00  6.714e-01   5.313  1.11e-07 ***
wd         -1.390e-01  6.018e-03 -23.094  < 2e-16 ***
rad        -7.337e-03  3.157e-03  -2.324  0.020136 *

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 29.78 on 7271 degrees of freedom  
Multiple R-squared: 0.5598, Adjusted R-squared: 0.5592

F-statistic: 924.7 on 10 and 7271 DF, p-value: < 2.2e-16

### Estación PPD – Modelo II - Validación PM<sub>10</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm10_PPD)
```

Residuals:

Min	1Q	Median	3Q	Max
-70.808	-15.937	-0.553	14.493	79.096

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	47.65541	1.34765	35.36	<2e-16	***
observado	0.55453	0.01161	47.78	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.24 on 1819 degrees of freedom  
Multiple R-squared: 0.5565, Adjusted R-squared: 0.5563

F-statistic: 2283 on 1 and 1819 DF, p-value: < 2.2e-16

### Estación VMT – Modelo I - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = VMT ~ SBJ + CDM + HCH + SJL + SMP + PPD, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-36.719	-6.170	-0.328	6.070	38.749

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.43238	1.04895	4.226	2.60e-05	***
SBJ	0.12344	0.04604	2.681	0.00745	**
CDM	0.13287	0.05277	2.518	0.01195	*
HCH	0.12349	0.02737	4.511	7.20e-06	***
SJL	0.12940	0.03282	3.943	8.60e-05	***
SMP	0.16678	0.05630	2.962	0.00312	**
PPD	0.23283	0.03275	7.109	2.19e-12	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.743 on 1018 degrees of freedom  
Multiple R-squared: 0.3907, Adjusted R-squared: 0.3871

F-statistic: 108.8 on 6 and 1018 DF, p-value: < 2.2e-16



### Estación VMT – Modelo I - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_VMT)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-13.0537	-3.7268	-0.3784	3.3770	19.3055

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	15.94513	0.86557	18.42	<2e-16	***
observado	0.40363	0.02869	14.07	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.531 on 254 degrees of freedom

Multiple R-squared: 0.438, Adjusted R-squared: 0.4358

F-statistic: 197.9 on 1 and 254 DF, p-value: < 2.2e-16

### Estación VMT – Modelo II - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = pm25 ~ pm10 + co + no2 + so2 + temp + HR + ws +  
wd + rad, data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-31.725	-6.166	-1.068	5.136	46.395

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.431e+01	3.388e+00	-4.224	2.44e-05	***
pm10	6.549e-02	2.353e-03	27.840	< 2e-16	***
co	6.505e-03	5.318e-04	12.232	< 2e-16	***
no2	1.435e-01	1.604e-02	8.943	< 2e-16	***
so2	7.518e-01	3.714e-02	20.241	< 2e-16	***
temp	2.046e-01	6.891e-02	2.969	0.003	**
HR	1.741e-01	2.537e-02	6.864	7.44e-12	***
ws	-5.332e+00	2.370e-01	-22.493	< 2e-16	***
wd	2.363e-02	2.283e-03	10.350	< 2e-16	***
rad	1.807e-02	6.649e-04	27.182	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.408 on 5592 degrees of freedom

Multiple R-squared: 0.487, Adjusted R-squared: 0.4862

F-statistic: 589.9 on 9 and 5592 DF, p-value: < 2.2e-16

### Estación VMT – Modelo II - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_VMT)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-20.2474	-4.7211	-0.6929	3.9457	24.2204

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.00061	0.38163	36.69	<2e-16 ***
observado	0.45530	0.01353	33.66	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.701 on 1398 degrees of freedom

Multiple R-squared: 0.4476, Adjusted R-squared: 0.4473

F-statistic: 1133 on 1 and 1398 DF, p-value: < 2.2e-16

### Estación CDM – Modelo I - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = CDM ~ SBJ + VMT + HCH + SMP, data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-20.728	-3.754	-0.143	3.714	19.522

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.91259	0.52957	5.500	4.80e-08 ***
SBJ	0.21612	0.02640	8.185	8.06e-16 ***
VMT	0.04865	0.01778	2.737	0.00631 **
HCH	0.06257	0.01438	4.350	1.50e-05 ***
SMP	0.37237	0.03028	12.297	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.768 on 1020 degrees of freedom

Multiple R-squared: 0.4538, Adjusted R-squared: 0.4517

F-statistic: 211.9 on 4 and 1020 DF, p-value: < 2.2e-16

### Estación CDM – Modelo I - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_CDM)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-10.1669	-2.4536	-0.3749	2.1777	12.5749

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	8.75884	0.56389	15.53	<2e-16	***
observado	0.45193	0.03184	14.19	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.794 on 254 degrees of freedom

Multiple R-squared: 0.4423, Adjusted R-squared: 0.4401

F-statistic: 201.5 on 1 and 254 DF, p-value: < 2.2e-16

### Estación CDM – Modelo II - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = pm25 ~ pm10 + co + no2 + so2 + o3 + HR, data = training)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.7446	-2.6352	-0.4574	2.1922	16.2828

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.9427967	0.7671689	-3.836	0.000127	***
pm10	0.2051170	0.0048339	42.433	< 2e-16	***
co	0.0043164	0.0005351	8.067	9.11e-16	***
no2	-0.0985126	0.0107489	-9.165	< 2e-16	***
so2	-0.1347875	0.0255593	-5.274	1.40e-07	***
o3	0.0431948	0.0060148	7.181	8.02e-13	***
HR	0.1057327	0.0080528	13.130	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.075 on 4551 degrees of freedom

Multiple R-squared: 0.3767, Adjusted R-squared: 0.3759

F-statistic: 458.5 on 6 and 4551 DF, p-value: < 2.2e-16

### Estación CDM – Modelo II - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_CDM)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.9213	-1.6788	-0.2538	1.4504	9.5225

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	7.45084	0.19484	38.24	<2e-16	***
observado	0.38780	0.01459	26.58	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.515 on 1138 degrees of freedom  
Multiple R-squared: 0.3831, Adjusted R-squared: 0.3826

F-statistic: 706.7 on 1 and 1138 DF, p-value: < 2.2e-16

### Estación PPD - Modelo I - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = PPD ~ ATE + VMT + HCH + SJL + SMP + CRB, data =  
training)
```

Residuals:

Min	1Q	Median	3Q	Max
-34.901	-5.263	-0.201	5.291	34.379

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	10.79610	0.85215	12.669	< 2e-16	***
ATE	0.10724	0.02547	4.210	2.78e-05	***
VMT	0.19083	0.02698	7.072	2.83e-12	***
HCH	0.08799	0.02641	3.331	0.000896	***
SJL	0.08885	0.03073	2.892	0.003914	**
SMP	0.16734	0.04484	3.732	0.000201	***
CRB	0.24489	0.02734	8.958	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.664 on 1018 degrees of freedom  
Multiple R-squared: 0.5041, Adjusted R-squared: 0.5011

F-statistic: 172.4 on 6 and 1018 DF, p-value: < 2.2e-16

### Estación PPD - Modelo I - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_PPD)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.924	-4.965	-1.131	4.076	22.146

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	19.988	1.265	15.79	<2e-16	***
observado	0.442	0.035	12.63	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.62 on 254 degrees of freedom  
Multiple R-squared: 0.3857, Adjusted R-squared: 0.3833

F-statistic: 159.5 on 1 and 254 DF, p-value: < 2.2e-16

### Estación PPD – Modelo II - Entrenamiento PM<sub>2.5</sub>

Call:

```
lm(formula = pm25 ~ ., data = training)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-36.625	-6.252	-0.717	5.675	39.148

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-4.635e+01	3.015e+00	-15.374	< 2e-16	***
pm10	1.592e-01	3.097e-03	51.413	< 2e-16	***
co	9.140e-04	2.774e-04	3.295	0.000989	***
no2	8.551e-02	7.978e-03	10.718	< 2e-16	***
so2	1.220e-01	1.079e-02	11.307	< 2e-16	***
o3	9.196e-02	1.433e-02	6.417	1.48e-10	***
temp	2.452e-01	5.975e-02	4.104	4.10e-05	***
HR	6.320e-01	2.248e-02	28.108	< 2e-16	***
ws	-4.736e+00	1.999e-01	-23.698	< 2e-16	***
wd	1.792e-02	1.911e-03	9.381	< 2e-16	***
rad	2.328e-02	9.346e-04	24.913	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.184 on 7271 degrees of freedom

Multiple R-squared: 0.458, Adjusted R-squared: 0.4572

F-statistic: 614.4 on 10 and 7271 DF, p-value: < 2.2e-16

### Estación PPD – Modelo II - Validación PM<sub>2.5</sub>

Call:

```
lm(formula = pronosticado ~ observado, data = pm25_PPD)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-21.8979	-3.9882	-0.3932	3.8229	24.7979

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	17.3786	0.3862	45.00	<2e-16	***
observado	0.4407	0.0115	38.32	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.205 on 1819 degrees of freedom

Multiple R-squared: 0.4466, Adjusted R-squared: 0.4463

F-statistic: 1468 on 1 and 1819 DF, p-value: < 2.2e-16